THE AIR FORCE OPERATIONAL RISK MANAGEMENT PROGRAM AND AVIATION SAFETY

THESIS

Matthew G. Cho, Captain, USAF

AFIT/GLM/ENS/03-02

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GLM/ENS/03-02

THE AIR FORCE OPERATIONAL RISK MANAGEMENT PROGRAM AND AVIATION SAFETY

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Matthew G. Cho, BArch

Captain, USAF

March 2003

AFIT/GLM/ENS/03-02

THE AIR FORCE OPERATIONAL RISK MANAGEMENT PROGRAM AND AVIATION SAFETY

Matthew G. Cho, BArch
Captain, USAF

Approved:

____________________________________ ____________
Stephen M. Swartz, Lt Col (USAF) (Advisor) date

____________________________________ ____________
Stanley E. Griffis, Maj (USAF) (Reader) date

Acknowledgments

I would like to express my sincere appreciation to my faculty advisor, Lt. Col. Stephen Swartz, and my reader, Maj. Stan Griffis, for their guidance and support throughout the course of this thesis effort. The insight and experience was certainly appreciated. I would also like to thank the personnel of the Air Force Safety Center for both the support and latitude provided to me in this endeavor. Most importantly, I am deeply indebted to my classmates, friends, and family for all the support, friendship, and love they provided me over the last eighteen months. I truly could not have done it without them.

Matthew G. Cho

Table of Contents

List of Figures

List of Tables

AFIT/GLM/ENS/03-02

Abstract

Aviation mishaps are extremely costly in terms of dollar value, public opinion, and human life. The Air Force drastically reduced Class A mishap rates in its formative years. The rate plummeted from 44.22 mishaps per 100,000 flight hours in 1947 to 2.33 mishaps in 1983 and has held steady around 1.5 mishaps since. The Air Force implemented the Operational Risk Management (ORM) program in 1996 in an effort to protect their most valuable resources: aircraft and aviators. An AFIT thesis conducted in 1999 by Capt Park Ashley studied the Army's similar Risk Management (RM) program. Ashley concluded that since his analysis found that RM did not affect the Army's mishap rates, the AF should not expect to see its rates decline due to ORM implementation.

The purpose of this thesis was to determine whether the implementation of ORM has had any affect on the AF's mishap rates. Analysis was conducted on annual and quarterly mishap rates, quarterly sortie mishap rates, and individual mishap data using three statistical techniques: comparison of means testing, discontinuous piecewise linear regression, and chi-squared goodness of fit testing. Results showed that the implementation of ORM did not effectively reduce the Air Force's aviation mishap rates.

THE AIR FORCE OPERATIONAL RISK MANAGEMENT PROGRAM AND AVIATION SAFETY

I. Introduction

Background

Man's quest to fly has always been accompanied by mishaps that take lives, destroy or damage aircraft, and cost countless dollars in damages. Although technology and experience have made flying a much safer endeavor, the inevitable losses are staggering. Military aircraft are particularly susceptible to mishap, given the combat role of many military airframes. Since its birth in 1947, the Air Force has lost 6,849 pilots and 13,626 aircraft, both of which are the Air Force's most precious resources (AF Safety Center, 2002). Despite the drastic reduction in mishap rate, between 1990 and 1996 the Department of Defense (DoD) suffered aviation losses of over $9.4 billion (Department of Defense, 1997).

Given the importance of these resources, improving aviation safety is critical. Traditional measures of mishap prevention are aircraft technological improvements and flight mishap investigations. Because human error contributes to the majority of aviation mishaps and is a contributing factor to approximately 70 percent of DoD Class A mishaps (Air Force Safety Center, 2003b), another methodology which focused on the aviator was needed. A study conducted by the Defense Science Board Task Force on Aviation Safety concluded that initiating a program of risk management for all the services would be the most efficient and effective means of reducing mishaps (Department of Defense, 1997).

The Army began fielding a risk management program formally in 1987 and enjoyed a reduction in its Class A mishap rate since. The Air Force Operational Risk Management (ORM) program was implemented in Sep 1996 as a means to reduce aviation mishaps. The program was intended to enhance safety and overall mission effectiveness by instilling a structured system of decision-making processes to evaluate situations, identify risks, and determine optimal courses of actions.

Air Force leadership recently indicated that they were moderately pleased with the progress of the ORM program thus far, but were looking for improvements in the future. General John P. Jumper, Air Force Chief of Staff, upon reviewing the results from an Inspector General ORM Eagle Look in early 2002, released a memorandum in June addressing the program status (Jumper, 2002a).

According to the memorandum, the Air Force had been moderately successful in the implementation of the program goals, but was not as far along as it could be. General Jumper cited the Eagle Look as reporting a general lack of leadership emphasis and inadequate training programs as the primary areas of improvement. The memorandum called for senior leaders and commanders to put a higher priority on ORM, noting that the program cannot reach its maturity without their improved participation. Additionally, General Jumper directed leaders and commanders to emphasize training and to remain active in the overall ORM process (Jumper, 2002a).

Captain Park Ashley conducted a thesis (Ashley, 1999) on the Risk Management (RM) program used by the Army. His objective was to develop a predictive tool to estimate the future success of the Air Force ORM program. His work showed that RM did not improve the Army's mishap rates, and raised questions as to the potential efficacy

of ORM as an accident preventive treatment for the Air Force. Enough time has now passed with the Air Force experience to perform a more thorough study to determine whether the ORM program has been successful.

Problem Statement

Aviation mishaps are extremely costly in terms of dollar value, public opinion, and human life. The Air Force drastically reduced Class A mishap rates in its formative years. The rate plummeted from 44.22 mishaps per 100,000 flight hours in 1947 to 2.33 mishaps in 1983 and has held steady around 1.5 mishaps since (Air Force Safety Center, 2002). In an effort to protect their most valuable resources, aircraft and aviators, by further reducing modern mishap rates, the Air Force implemented the ORM program in 1996, designed to establish an atmosphere of safety at all levels.

Because recent studies of the Army's RM program, the model for the Air Force's ORM program, revealed that the program did not significantly improve Army aviation mishap rates, despite previous claims. In fact, evidence was found suggesting that accident rates actually increased during RM implementation. The study concluded that the Air Force should therefore not expect mishap rates to decline due to implementation of the ORM program.

Research Question

To what degree has the implementation of ORM affected flying safety in the Air Force?

Investigative Questions

The objective of this thesis effort is to analyze the efficacy of the ORM program in the reduction of aviation mishaps by tracking mishaps rates before, during, and after

ORM implementation. Known causal factors will be investigated as well in an effort to determine the contribution of ORM to mishaps. This research hopes to assist the Air Force effort to create a safer, more efficient organization. The following investigative questions (IQ) will be addressed and answered in the proceeding chapters.

IQ.1. What factors are involved in an aviation mishap?

IQ.2. What is ORM and how is it applied and implemented?

IQ.3. Have mishaps rates changed significantly since ORM was implemented?

IQ.4. Are any differences caused by ORM?

IQ.5. Have the proportion of human factors mishaps changed since implementation of ORM?

Methodology

The Chapter 2 Literature Review addresses investigative questions 1 and 2 which are qualitative in nature and best answered by a thorough review of Air Force policy, mishap journals, documents and texts, and other Department of Defense (DoD) safety literature.

To answer investigative questions 3, 4, and 5, a quantitative, statistical analysis of historical Air Force and Army mishap data was conducted. Several methods of analysis and time series techniques were used and are discussed in Chapter 3, Methodology.

Data Source

AF aviation data was gathered from the Air Force Safety Center (AFSC), Kirtland AFB, New Mexico. Annual mishap rates and mishap numbers are available online at the AFSC website and includes Class A, B, and C mishap numbers and rates from 1947 to

2001. Army aviation data was obtained from the Army Safety Center (ASC). Additional causal data were provided upon request by AFSC and ASC.

Scope and Limitations

The focus of this thesis will be Air Force aviation mishaps. This effort will study primarily Class A aviation mishaps: those that cost more than one million dollars, destroy an aircraft, or result in the loss of a life. Less catastrophic Class B data will also be analyzed to determine what additional effects ORM may have had. Army Class A, B and C data was also analyzed to confirm Ashley's findings.

Statistical procedures for non-parametric data differ from that of parametric data. Where the delineation between the two types of data was unclear, both types of procedures were used for the sake of thoroughness.

Assumptions

Significant non-compliance with ORM regulations would change the results of this study, however, determining whether personnel are actually utilizing ORM tools and instructions is another field of study that has not been addressed. Therefore, this thesis assumes that personnel are adhering to Air Force and Army directed implementation of the ORM program.

It is imperative to note that the implementation of an organization-wide effort such as ORM does not happen instantaneously. The Air Force officially began its ORM program on 2 Sep 96, but full implementation, accomplished via individual computer awareness training was not completed until 1 Oct 98. This potentially confounding two-year implementation period was accounted for in analyses in Chapter 4.

Summary

This chapter introduced the Air Force ORM program and identified the objective of determining its affect on aviation safety. It discussed the background, problem, investigative questions, methodology, data source, scope, and assumptions of this thesis document. The next four chapters of this research effort include the Literature Review, Methodology, Analysis and Results, and Conclusions.

The Literature Review provides a broad overview of the nature of aviation mishaps, the Air Force and Army risk management programs, and other issues relevant to the research objective. The findings contained within were essential to defining the scope of the project, developing an understanding of the subject matter, and laying foundations for the statistical analysis of the mishap data.

The Methodology chapter describes the various statistical methods, tests, and techniques used to analyze the data. It also details the typology of the research design and the various threats to validity.

The Findings and Analysis chapter presents the data obtained and the results of the statistical analysis. This section answers the investigative questions posited in Chapter 1 and discusses the end results of the research effort.

The Conclusions chapter ends the thesis by presenting the research findings and their relevance and significance. This chapter also poses recommendations for the future and potential topics for future study in the arena of aviation mishaps.

II. Literature Review

Overview

The goal of this literature review is to provide a background into the various aspects of aviation safety and its relationship to operational risk management. Initially, the various aviation safety factors are identified and described and mishap prevention methods are discussed. A discussion of relevant risk management and safety terms and definitions as defined by the Air Force and Army are then provided. Finally, the implementation of risk management by both the Air Force and Army is outlined.

Aviation Safety Factors

There are countless numbers of factors that affect aviation safety: bird strikes, fatigue, weather, psychological conditions, parts failure, controlled flight into terrain, operations tempo, etc. Ashley (1999) identified a model that incorporates these factors. The model is shown in Figure 1.

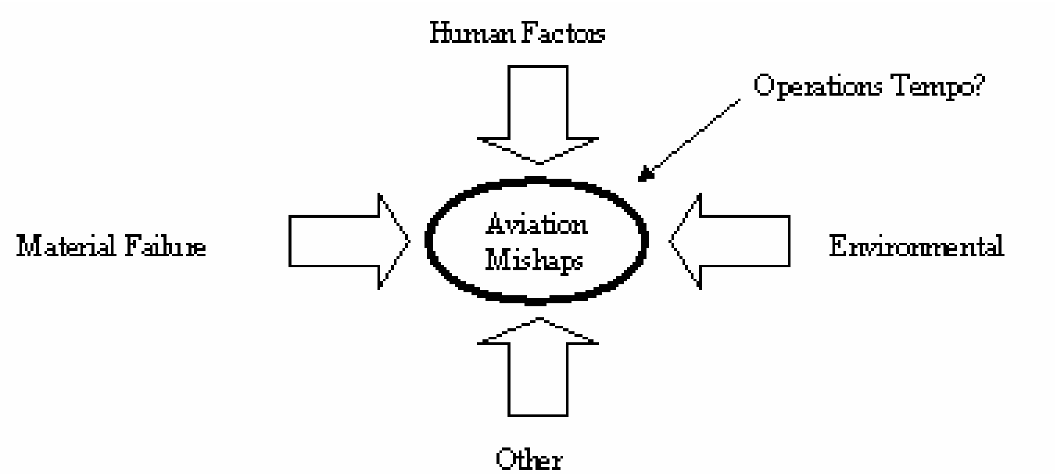


Figure 1. Aviation Mishap Cause Factors (Ashley, 1999)

The model follows the four mishap cause classifications outlined in DODI 6055.7; human factors, material failure, environmental, and other (Department of Defense, 1989). Both the Air Force and the Army follow this basic model for the purposes of classifying mishap causes. Ashley also identified a fifth possible factor--operations tempo. These five primary cause factors will now be discussed.

Human Factors.

Human factors describe mishap causes that relate to human error or the human condition. Primarily, they refer to the pilot of the flown aircraft but the factors may also pertain to involved ground crew and supervisory roles. Examples of human factors include poor judgment, improper risk assessment, psychological conditions, physiological conditions, and many more. All of which, alone or in conjunction with other factors, can lead to an aviation mishap. Several key human factors related concepts are now discussed in greater detail, including a classification system, age, and controlled flight into terrain.

The Human Factors Analysis and Classification System (HFACS).

Due to the high rate of mishaps attributed to human factors (between 60 and 80%) much research has been conducted on the causes of human error. Studies of specific failures in human decision making led to the development of HFACS (Shappell and Wiegmann, 2000). HFACS is a tool used to identify and classify the human factors causing aviation mishaps and is employed by all of the services in aviation accident investigation and analysis.

HFACS is based on the premise that human factor aviation accidents are not isolated incidents; rather, they are the result of a definite chain of events that lead to

unsafe aircrew behavior and ultimately, an accident. HFACS is used to assist accident investigations in uncovering and categorizing the causes of mishaps and aid in the development of safer practices.

The system, which has been embraced by many in the aviation industry, defines four tiers of an accident's chain of events. They are first) organizational influences, second) unsafe operations, third) preconditions for unsafe acts, and fourth) the actual unsafe acts of the aircrew. The HFACS further delineated 17 causal categories of human errors within those four tiers.

In the first tier, Organizational Influences, improper resource management, unsafe organizational climates, or poor organizational processes are identified as possible causes of mishaps (Shappell and Wiegmann, 2000).

The second tier, Unsafe Supervision, refers to inadequate supervision, inappropriately planned operations, uncorrected problems, and supervisory violations. The first and second tier are only applicable in commercial and military environments, where organizations and leaders are involved in flying operations and are not applicable in general aviation, where aircraft are privately operated (Shappell and Wiegmann, 2000).

The third tier, Preconditions for Unsafe Acts, includes substandard conditions of the operator, such as adverse mental and physiological states and physical and mental limitations. Also included are substandard practices of the operators; either failures in crew resource management or personal readiness (Shappell and Wiegmann, 2000).

The final tier, the Unsafe Acts of the Operators, is comprised of violations, both routine and exceptional, and errors, including decision, skill-based, and perceptual errors (Shappell and Wiegmann, 2000).

Age.

One possible source of human errors in aviation mishaps is pilot age. A study was conducted in 2002 to determine whether pilots of different age groups believed that their piloting skills, such as reaction speed, concentration, and decision making had deteriorated over time. The study, which polled over 1,300 airline pilots, used questionnaires employing the 5-point Likert rating system to rate their abilities at the present and in the past. The results of the test showed that most pilots, regardless of age, reported that their abilities declined while under stress and anxiety. It also concluded that older pilots were not more likely than younger pilots to report negative changes in their abilities, suggesting that age is not perceived by aviators as a significant cause of error (Rebok and others, 2002).

Based on the literature, it is inconclusive whether age is a direct factor in aviation mishaps. It would seem more likely that physiological factors associated with aging would have a more profound affect. The mean age of aircrew involved in AF Class A and B mishaps was approximately 31 years of age. Unfortunately, since successful sortie data was not available, we cannot draw any conclusions about whether age has an impact on the likelihood of mishap occurrences.

CFIT.

Controlled Flight Into Terrain, or CFIT, occurs when an aircraft flies into either water or land due to the inadequate situational awareness of the pilot. It is a significant type of human factors related aviation mishap in the military, commercial, and general aviation environments. The Navy/Marine Corps lost an average of ten aircraft per year

due to CFIT between 1983 and 1995. Between 1990 and 1999, 32% of all commercial airline fatalities, adding up to over 2,100 deaths, occurred because of CFIT; the single greatest contributor to commercial losses. And in a two-year period between 1993 and 1994, the Federal Aviation Administration (FAA) identified 195 CFIT incidents (Shappel and Wiegmann, 2001).

In a study conducted by Shappell and Weigmann in 2001, it was determined that approximately 50% of CFIT mishaps were associated with decision errors, 45% with skill-based errors, 30% with violations, and 20% with perception errors. Their research, aided by the HFACS, also determined that the use of decision making aids and recurring pilot training would decrease the likelihood of CFIT incidents (Shappel and Wiegmann, 2001). Despite the significant number of CFIT incidents, it is not considered a cause factor in and of itself, but rather a combination of various human factors.

Material Causes.

The second largest causal contributor to aviation mishaps is material failure. From 1993 to 1998, the Air Force experienced material related mishaps in 12% of Class A accidents, 27% of Class B accidents, and 39% of Class C accidents (Ashley, 1999). Aircraft are composed of thousands of intricately interwoven complex parts. It is only natural then that failures occur. Although a material failure would likely be traced back to a human error at some point in its production life cycle, this thesis is studying the immediate causes of mishaps and so material failures remain an important topic. Material failures include faulty parts due to wear and tear and design and manufacturing problems. The Air Force recognizes faulty design, parts failure, and manufacturing failures as contributors to this mishap category. Similarly, the Army refers to instances

when materiel elements become inadequate as “Materiel Factors.” (Department of the Army, 1999)

Environmental Factors.

Another important are of mishap factors is environmental factors. Environmental factors include contributors such as weather and wildlife strikes. Aviation mishaps involving environmental mishaps are fairly common, with weather and bird strikes being the most common. It should be noted that many mishaps with environmental factors involved are not solely blamed on the environmental cause, but are instead identified in conjunction with other human factors involving the failure to avoid the environmental obstacle. Both the Air Force and the Army identify environmental factors as contributing factors to aviation mishaps.

Weather.

Adverse weather conditions cause accidents every year and are considered to be one of the major contributing factors to aviation mishaps. Weather conditions not only cause accidents outright but also contribute to mishaps caused by human factors. A study conducted at the Naval Postgraduate School determined that 12% of all Naval Class A mishaps between 1990 and 1998 were weather related and that a further 19% of human factors mishaps during the same time period were also weather related (Cantu, 2001). Furthermore, statistics from a study of commercial aviation conducted by the FAA concur, concluding that 12% of fatal U. S. commercial carrier accidents were directly caused by weather conditions (Duquette, 1998).

The weather is clearly a major contributor to aviation safety concerns. It is unpredictable and can be treacherous in numerous ways. Visibility and ceiling

conditions, including fog, low ceilings, clouds, obscurity, and sand storms are all dangerous factors that aviators must contend with. The wind is also a dangerous element. Crosswind, tailwind, gusts, and wind shear all contributed to accidents in Cantu's study. Furthermore, the environment can produce icing problems, turbulence, precipitation, and electrostatic discharges that can adversely affect safe flying operations. The major sources of adverse weather conditions were poor visibility (54%), wind (16%), and precipitation (12%) (Cantu, 2001).

Bird Strikes.

Contributing to the environmental dangers of aviation are the populations of birds taking residence near airports. Despite their diminutive size relative to aircraft, bird strikes are responsible for a considerable number of mishaps each year. Typically, such mishaps are caused when birds, many of which are endangered species and cannot be exterminated, are ingested into engine intakes, causing immediate damage and forcing engine failure. Additionally, the dangers of direct impact are also considerable. According to one study, a twelve-pound fowl struck by an aircraft traveling at 150 mph generates the force of a thousand pound weight dropped from a height of ten feet (Birdstrike Committee USA, 2002). Since 1973, the Air Force has suffered 32 aircraft losses and 35 fatalities due to bird strikes. In an effort to reduce such numbers, the Air Force created the Bird/Wildlife Aircraft Strike Hazard Team to study the phenomena and work to solve the bird strike problem (BASH, 2002).

Operations Tempo.

In March 1999, two HH-60G Pave Hawk helicopters based at Nellis AFB, Nevada collided in mid-air, killing all twelve crewmembers aboard. The ensuing

accident investigation report indicated that an unrelenting operations tempo was the underlying cause for the aircrew errors that caused the accident. The squadron had recently been engaged in two simultaneous deployments and had been home only 10 months out of the previous 3 years (Brandon, 1999). Clearly, high operations tempo can be a contributor to aviation mishaps.

Operations Tempo is a term widely used in discussions of today's military forces. It generally refers to the workload of both organizations and individuals and is generally seen as an impediment to readiness and performance. The Air Force defines operations tempo as the sum total of all activities a unit is involved in. It includes deployments, TDY, inspections, productivity days, extended workdays, and normal workdays. Due to recent awareness of high operations tempo, legislation has been developed forcing the services to more closely define tempo and more accurately track and compensate individual hardships.

Military leadership seems to agree that operations tempo is at an all time high. Two years before the events of September 11, 2001 and the subsequent actions in Afghanistan, all four services testified before the Senate Armed Services Committee that operations tempo was a major problem. General Ryan of the Air Force reported that despite a force 40% smaller than during the Cold War, the Air Force was deploying four times as often. General Shinseki of the Army testified that his service was busier than he had ever seen in his 35 years of experience. All representatives agreed that smaller force structure combined with greater demands and insufficient budgets were creating problems (Status of the United States Military, 2002). Continued operations since in Kosovo, East Timor, and the Middle East have added to the workload.

Ashley (1999) identified Operations Tempo as a possible category of mishap factors, along with human, environmental, and material factors. This literature review of operations tempo draws the conclusion that it is not a major category of mishap factors. Instead, operations tempo, much like CFIT, is a combination of a number of other factors, including organizational and physiological affects.

It remains unclear to what degree operations tempo affects safe flying operations. In his thesis, Ashley (1999) notes that two separate studies, one conducted by the Air Force in 1994 and one by a Blue Ribbon Panel in 1995, found no direct statistical correlation between operations tempo and aviation mishaps. Nevertheless, sustained periods of high operations tempo are often associated with psychological stress, fatigue, and emotional duress; a combination of human factors mishap contributors. Recent studies have indicated that operations tempo can be linked to problems with retention, family stability, and medical readiness; all of which could be contributors to piloting or even maintenance errors (Castro & Adler, 1999).

The next section of the literature review describes the differences between the Air Force and Army systems of mishap cause factor identification.

Air Force Cause Factors

When determining the cause of an aviation mishap, the Air Force investigating agent first identifies a person or functional area as the causal finding agent. Then a causal finding area is identified. These areas are broadly defined categories within which the mishap occurred and include Logistics, Maintenance, Environmental, Operations, Support, and Unknown. These categories and detailed explanations follow, and are

found in AFI 91-204, Safety Investigation and Reports (Department of the Air Force, 2001).

The Logistics area refers to findings related to acquisition, manufacturing, design, and procurement that do not involve individual maintenance or operations personnel. The Maintenance category attributes the cause to AF or contracted maintenance personnel. Environmental factors are causal findings relating to animals or environmental conditions that could not be reasonably avoided. The Operations area refers to the actual aviators involved. Support areas include the various support functions at installations, including Civil Engineering, Supply, Transportation, etc.

Once a general causal finding area is designated, the investigators determine the specific reasons for the occurrence of the causal finding. These reasons are categorized into four distinct areas: People, Parts/Paper, Natural Phenomena, and Unknown.

People.

People reasons are related directly to individuals involved in the finding and is further divided into three areas; Physical, Personnel, and Psychological Reasons. Physical Reasons refer to factors affecting the individual's body and state of wellness. Factors include:

- Ergonomic considerations: weight or strength
- Self Induced Stressors: voluntary medication usage and alcohol abuse
- Pathological: mental or emotional illness
- Perceptions: misinterpretations of the environment and failure to react to surroundings

- Physiological: problems or adverse conditions caused by normal biological functions, such as hyperventilation and fatigue

Personnel Reasons are based on the qualifications of the individual involved in the mishap, including proficiency, manning, training, and unauthorized modifications. Proficiency reasons arise when individuals were properly trained and qualified at one time, but lacked the skills at the time of the incident to perform adequately. Manning reasons occur when there are not enough qualified personnel available to properly accomplish the event. Training reasons refer to situations where individuals are not sufficiently trained for the task. Unauthorized modifications are modifications to equipment and/or aircraft made without official approval.

Psychological Reasons refer to cognitive decisions and functions made by the causal individual. Acceptable reasons include:

- Accepted Risk: a risk assessment was conducted correctly
- Attention Management: distractions and inattention
- Cognitive Function: misinterpretation of data, insufficient aptitude
- Discipline: intentional non-compliance of standards, “horseplay”
- Emotional State: personal feelings resulting in adverse behavior such as moodiness, complacency, and over-motivation
- Inadequate Risk Assessment: actions were taken without conducting a sufficient risk assessment
- Judgment: poor assessment of information
- Preparation: inadequate mission planning factors

Parts/Paper.

Parts/Paper Reasons are attributed to faulty equipment, design, or publications that contribute to the mishap. This category of mishap reasons is further subdivided into five categories: attrition, design, faulty-part, manufacture, and publications. Attrition refers to "decisions made to replace by attrition rather in lieu of issuing a time compliance technical order or retrofit package." (Department of the Air Force, 2001) Design reasons occur when systems are inadequately designed or do not meet specifications. Faulty-parts include defective aircraft parts or personal equipment. Manufacture reasons are designated when equipment is defective due to manufacturing problems. Publication reasons occur when technical orders, instructions and other paper work are misleading or inadequate.

Environmental.

The Air Force categorizes natural phenomena reasons into either Animal or Environmental Conditions categories. Animal reasons refer to collisions with animals, ingestion of animals into engine intakes, and mishaps due to attempted avoidance of animals. Environmental Conditions occur due to weather conditions such as high winds and fog. It should be noted that these occurrences are used as reasons only when reasonable preparations were made to avoid them. For example, a lightning strike during a thunderstorm cannot be the cause of a mishap if an adequate weather forecast was available and the thunderstorm could have been avoided.

Unknown.

Any mishap for which a reason cannot be determined falls into the Unknown category. Investigators who categorize mishaps into this category must provide a fully descriptive narrative explaining the selection of this reason.

Army Causes

The Army recognizes three categories of aviation mishap causes; human factors, materiel factors, and environmental factors. Army Regulation 385-40 describes procedures for the investigation and classification of the mishaps into the three broad categories, but are not as specific as the Air Force in labeling the causes.

Human Factors.

The Army defines Human Factors as "human interactions (man, machine, and/or environment) in a sequence of events that were influenced by, or the lack of human activity, which resulted or could result in an Army accident." Human factors that lead to accidents are largely caused by human error, which is also defined by the regulation. Human errors are human acts that deviated from the operational requirements of the act. The reporting of human factors as a causal agent in Army mishaps is less categorized than the Air Force reporting system, but according to the regulation, such errors are generally attributable to inefficiencies in training, standards, leaders, the individual, or other support areas.

Materiel Factors.

Materiel factors may cause accidents when materiel elements of a vehicle, equipment, or a system become inadequate or counter-productive to its operation. This includes design failures and malfunctions that directly lead to an aviation mishap.

Environmental Factors.

Environmental Factors may be the cause of a mishap when environmental conditions, such as weather or wildlife, could have adversely affected the equipment or aircraft being operated or the actual performance of the individual operators.

Prevention Factors

The ultimate goal of identifying mishap causes is to develop a prescription to prevent their future occurrence. This section addresses areas of mishap prevention including human factors programs, mishap investigation, leadership, and technological advancements.

Human Factors Programs.

One of the aviation industry's efforts to reduce aviation mishaps due to human factors was the implementation of Crew Resource Management (CRM). CRM is a training system that was developed to enhance the interaction between members of a flight crew. It consists of training focused on standardized procedures and operations that sharpen aircrew communications. Crews are trained to work together more fluidly and efficiently by task sharing during high workload scenarios. Such methods were designed to create a more effective sense of teamwork and to subsequently reduce miscommunication and mishaps. The Air Force instituted a CRM training program in 1994 and was found to be beneficial in multiple crew environments (Department of the Air Force, 1995).

CRM's effectiveness in reducing mishaps is questionable as other sources showed no statistical improvement due to CRM. Some studies have shown that while CRM training does effectively promote better communication patterns, learning, and team-

oriented behavior, its implementation has not significantly reduced mishap rates (Johnson, 2002) or that the results are indeterminate (Burke, et al., 2002). Some suggest that CRM's apparent failure occurs because its practices breakdown during critical or abnormal conditions, such as when aircrew are fatigued, when under heavy workload, or while subjected to dangerous environmental surroundings (Johnson, 2002).

HFACS, discussed earlier, is another key tool used to prevent human factors mishaps. HFACS assists investigators in the identification of mishap causes involving human factors, ultimately assisting in the dissemination of prevention information.

Mishap Investigations.

One of the Department of Defense's primary methods of accident prevention is the identification and understanding of past failures through accident investigations. By analyzing the causal factors involved in aviation accidents, better controls for the prevention of future ones can be developed and implemented. Generally, safety investigation boards are convened to collect evidence, analyze the data, and determine what caused the mishap. The board writes reports and disseminates the findings to aircraft designers and users to facilitate future prevention.

The DoD accident investigation system is based on four principles. First, investigations are carried out by an unbiased and disinterested third party to ensure a fair and even assessment of the situation. Assignment of investigative duties is prescribed by the services' instructions on accident investigation. Second, investigators will be assigned based on the necessary training and skills, commensurate with the severity and complexity of the investigation. Third, reports of the investigation are reviewed by the chain of command above the organization involved. Fourth, lessons learned are

disseminated to the entire community to aid future prevention and corrective actions are implemented and tracked (Schilder, 2002).

The investigation is a lengthy process. The accident is classified according to its severity. Data is collected to determine damages and injuries. The accident is categorized based on the causal factors and activities involved.

Once the investigation has been concluded, it essential that any lessons learned from the accident are disseminated to the involved community. There are many tools provided by the DoD to aid in this effort, including on-line safety databases and software. The Air Force, Army, and Naval Safety Centers are the focal points for safety discussions and the lesson dissemination.

Each service has its own process and agency responsible for investigations. This concept was challenged by a Defense Task Force study in 1997 to determine if there was a need for a joint agency. The task force noted that most investigations are rapidly accomplished within 60 days, have thorough participation by their associated Safety Centers, and are service specific based on mission needs. It was furthermore noted that although the services use separate procedures, they all successfully follow the fundamentals of accident investigation. The task force concluded that no joint office of investigation was needed (Department of Defense, 1997).

The Air Force investigation procedures are documented in AFI 91-204, Safety Investigations and Reports. This enormous, 519-page instruction outlines the entire process, beginning with the determination of responsibilities and the composition of the Safety Investigation Board (SIB), which investigates the accident. The SIB is responsible for all aspects of the investigation, including categorization of causal factors and

classification of the mishap. They conduct the investigation, which consists of data collection, evidence gathering, interviews and other procedures. The SIB prepares safety messages, media information releases, and a number of formal reports to document the investigation.

The Army conducts its investigations using the "3W" approach laid out in DA PAM 385-40, Army Accident Investigation and Reporting. First, the investigators determine what happened by identifying the causal factors; human, materiel, or environmental. Second, why it happened is investigated by examining the leadership, training, procedures, support, and the individual involved. Thirdly, what to do about it in the future is determined by making recommendations for fixes, remedial measures, and countermeasures.

The actual investigation takes place in phases. Phase I is the organization of the investigation board and a preliminary examination. Phase II is the data collection phase, where evidence is collected and grouped into the human, materiel, and environmental factor categories. Phase III sees data analyzed and findings compiled. Phase IV is the completion of the technical report for the recording of the investigations findings.

Policy for civilian investigations is developed by FAA. The guidelines for aircraft mishap investigations are documented in Order 8020.11B, Aircraft Accident and Incident Notification, Investigation, and Reporting (Department of Transportation, 2000). The National Transformation Safety Board (NTSB) is an outside agency that conducts the investigations. It coordinates throughout with the FAA, determining the causes of the incident and ultimately making recommendations for future prevention efforts.

Leadership.

The leadership role is one of the critical areas where ORM principles can be effectively utilized to improve safety. Both the Air Force and the Army identify leaders as shouldering the burden of ensuring ORM is adhered to under their command.

The Air Force states in its ORM policy directive that "all Air Force Personnel will apply ORM principles, concepts and techniques." It is an all-encompassing program for all ranks to adhere to. The role of the commander is to tailor the ORM program to the unit's needs and ensure the unit's implementation and sustainment of ORM into decision-making (Department of the Air Force, 2000c). The commander assumes the ultimate responsibility for ORM adherence at his or her level of command.

Air Force leadership is highly involved beginning at the headquarters level. AF Chief of Staff, General John P. Jumper, stated in his memo regarding ORM that "Air Force senior leaders and commanders at all levels have to provide the continuing emphasis necessary for ORM to reach full maturity." (Jumper, 2002a) The Chief of Staff establishes safety policy and guidance. Headquarters staffs serve as the principle advocates for the program, ensuring plans and programs are distributed and adhered to. The AF Safety Center serves as the lead agency and overall program manager for the integration of ORM into the AF. It also monitors advancements in the safety realm. Safety staffs are formed at all levels where needed, including MAJCOM, wing, and flight levels. Additionally, Flight Safety Officers and Functional Program Managers serve their commanders to ensure ORM is integrated, trained, and utilized throughout (Department of the Air Force, 2000c). A more detailed discussion of ORM responsibilities is included in a later section.

The Army field manual states that although the responsibility for safety runs throughout the ranks, “commanders—with the assistance of their leaders and staffs—manage accident risks,” and that only after the principles are embraced and enforced by commanders will the Army recognize “the full power of risk management.” (Department of the Army, 1998) While the aviator is identified as the core of aviation mishap prevention, leadership must be intimately involved.

Technological Improvements.

The drastic reduction in mishap rates since the early days of the AF in the 1940s are primarily attributed to technological improvements and innovations. General improvements in the aircraft, more reliable engines, and more sophisticated aviation systems have all contributed to the rate reductions. Technological improvements continue to advance the cause of safety, although their direct contributions to mishap reduction rates declines (Driskell and Adams, 1992).

Today, various organizations within the government are involved in finding safer ways to fly, including the FAA, NASA, DoD, and NTSB. Research is constantly being conducted to find better techniques and technologies in areas such as ejection seats, air control systems, weather forecasting systems, and aircraft component design. NASA is currently involved in several projects, including conducting research on fatigue and an aviation incident reporting system that would collect mishap data and dispense it via messages and alerts to its users. (Human Factors, 2002) The FAA is currently conducting high-tech weather studies to learn more about the effect weather has on flying (Aviation Studies, 2002).

Definitions and Concepts

The next section of the chapter is dedicated to defining the critical terminology involved in the ORM programs and aviation mishaps. The following terms are defined by DODI 6055.7, Accident Investigation, Reporting, and Record Keeping. This instruction outlines general procedures for reporting accidents to higher headquarters. It also contains a number of important definitions which are used by the DoD to standardize accident reporting.

Accident.

Accidents are defined by the DoD as an "unplanned event, or series of events, that results in damage to DoD property," (Department of Defense, 2000) and includes occupational illnesses and injuries to personnel. Also included under this definition are incidents whereby non-DoD property and individuals are damaged or injured due to DoD operations.

Accident Classification.

The DoD standardizes accident classifications using a system based on the severity of the mishap. Major accidents are designated as A, B, or C Class, with class A accidents being the most severe. Severity is determined by both the dollar value of lost assets (including environmental cleanup and restoration costs) and the resulting injuries or illnesses. Class C figures were changed in 2002 from a minimum value of $10,000 up to $20,000. Table 1 outlines DoD accident classifications.

Table 1. Accident Classification Specifications (Department of Defense, 2000)

Accident Class	Damage Costs	Injury
A	\$1,000,000 or more Destroyed Aircraft	Fatality Permanent Total Disability
B	\$200,000 < \$1,000,000	Permanent Partial Disability 3 or More Personnel Hospitalized
C	\$20,000 < \$200,000	Non-fatal Injury/Illness/Disability Causes Loss of Duty Time

Intent for Flight.

Intent for flight is an important term in the realm of aviation mishap categorization. It is used as a starting and stopping point for the classification of aviation accidents, as the existence of intent for flight differentiates between flight accidents and ground accidents. Intent for flight begins when an aircraft releases its brakes or when takeoff power is applied when beginning an authorized flight. Intent for flight ends when the aircraft has completed its flight and taxies clear of the runway. In the case of vertical landing aircraft, intent ends when the aircraft has touched down and is supported by its landing gear.

Types of Aviation Accidents.

The DoD categorizes all accidents into one of the following: aircraft, explosive and chemical agents, motor vehicles, ground and industrial, off-duty, unmanned aerial, guided missiles, maritime, nuclear, or space. Aircraft accidents are further segregated into three types of accidents: flight, flight-related, and ground accidents.

Flight accidents, which are the sole concern of this thesis, are accidents in which reportable damage to an aircraft occurred under circumstances where intent for flight

existed. Additionally, incidents involving explosives, missiles, or chemical agents causing damage to aircraft are reported in this category to avoid redundant reporting.

Flight-related accidents are accidents under intent for flight that occur when there is no damage to the aircraft, but involve injury or fatality to aircrew, ground crew, or passengers or damage to other property.

Ground accidents occur when there is injury or property damage without intent for flight, but while aircraft engines are in operation. This category includes flight decks of naval vessels.

Flight-related and ground accident occurrences are not used to calculate flight accident rates and are therefore not incorporated into this study.

Accident Rates.

Accident rates, which are commonly used to report aviation safety trends, are a measure of the recorded number of accidents per units of exposure. For example, Class A accident rates are calculated as the number of accidents per 100,000 flying hours.

Risk.

The fundamental concept in the development of risk management is that of risk itself. As a first step in eliminating aircraft accidents by reducing risk, one must understand what risk is. Risk, as a noun in this context, is defined as the "exposure to possible loss or injury (The Merriam Webster Dictionary, 1994)."

The military services define risk similarly while emphasizing intrinsically military aspects such as the role of 'adversaries' and 'personnel.' The Air Force definition is "an expression of consequences in terms of the probability of an event occurring, the severity of the event and the exposure of personnel or resources to potential loss or harm

(Department of the Air Force, 2000a)." The Army definition is "the probability and severity of a potential loss that may result from hazards due to the presence of an enemy, an adversary, or some other hazardous condition (Department of the Army, 1998)."

Risk Management.

The Air Force and the Army use slightly different terminology to describe the concept of risk management. The Air Force uses the term 'Operational Risk Management (ORM)' while the Army uses the term 'Risk Management (RM).'

The Air Force defines ORM as "a decision-making process to systematically evaluate possible courses of action, identify risks and benefits, and determine the best course of action for any given situation." (Department of the Air Force, 2000a) It calls for all levels to utilize the systems for all situations, both on- and off-duty. It states that proper implementation of the program will increase the overall strength of the Air Force's war fighting ability by enhancing mission accomplishment and preserving resources and individuals.

The Army defines RM as "the process of identifying, assessing, and controlling risks arising from operational factors and making decisions that balance risk costs with mission benefits." (Department of the Army, 1998) It calls for all levels, from soldiers to leaders, to implement risk management and states that the principles apply to all manners of operations and environments within the service. It notes the importance of leaders being able to properly apply risk management in order to conserve resources, protect personnel, and develop competent leaders and units.

ORM Principles.

AFI 90-901, Operational Risk Management (Department of the Air Force, 2000a), identifies four key principles that must be applied when managing risk. The instruction states that the principles should be used at all stages of decision-making; before, during, and after and should be used continuously when risk is present.

1) Accept no unnecessary risk: Unnecessary risk does not involve a positive return of benefits. From day to day operations to combat missions, risk is almost always involved at some level. The instruction urges members to accomplish their objectives while minimizing exposure to such unnecessary risks.
2) Make risk decisions at the appropriate level: Individuals who are accountable for the completion of operations are responsible for making risk decisions.
3) Accept risk when benefits outweigh the costs: Unlike situations involving unnecessary risk, acceptable risk involves gained benefits due to undertaken risk. Acceptable risk can be identified when potential costs are compared to potential benefits. Such undertakings are acceptable, but individuals should always attempt to minimize the risk and maximize the benefits.
4) Integrate ORM into operations and planning at all levels: From individuals at the lowest level conducting operations to the commanders at the highest levels establishing policy, risk management should be comprehensively applied.

The Army Risk Management program cites similar principles in FM 100-14, Risk Management (Department of the Army, 1998). Portions of the language used differ from

that of the Air Force's principles, but the message is generally the same. It identifies three key principles as a framework around which the program is built.

1) Integrating risk management into mission planning, preparations, and execution
2) Making risk decisions at the appropriate level in the chain of command
3) Accepting no unnecessary risk

Risk Management Process.

The Air Force's risk management process is based on several key fundamentals. It is designed to be a comprehensive system but must be tailored to fit each unique situation being addressed. The system outlines steps that provide tools for individuals to manage the immediate risk and provides six steps to use that define the risk management process.

1) Identify the Hazards: Use various hazard identification techniques to identify the hazards at hand.
2) Assess the Risk: The application of qualitative and/or quantitative assessment techniques should be used to determine the probability of risk and the implicit danger involved.
3) Analyze Risk Control Measures: Determine strategies that may be used to avoid, minimize, or eliminate the perceived risk.
4) Make Control Decisions: Make a decision at the appropriate level based on the cost-benefit analysis.
5) Implement Risk Controls: Carry out the selected strategy.

6) Supervise and Review: Continual review of the chosen strategy and the results of the action should be accomplished periodically to ensure success and improvement over time.

The Army incorporates a very similar five-step process while implementing its Risk Management program. The primary difference is that it combines control development and decision making into one step.

1) Identify Hazards
2) Assess Hazards to Determine Risks
3) Develop Controls and Make Risk Decisions
4) Implement Controls
5) Supervise and Evaluate

Fundamental Goals of ORM.

AFPD 90-9 highlights the overall goals of the ORM program (Department of the Air Force, 2000b). In general, understanding and minimizing risk will maximize mission effectiveness and ensure the highest levels of readiness for the Air Force. The directive identifies four separate goals:

1) Enhance mission effectiveness: Incorporation of ORM principles and practices will enhance all levels of mission effectiveness by preserving assets and keeping personnel safe and healthy.
2) Integrate ORM into processes: ORM should be integrated into mission processes at all times and decisions should be based on risk assessments.

3) Comprehensive acceptance at all levels: All personnel should be trained and motivated to use ORM in all situations where risk is involved, both on- and off-duty.
4) Improve war fighting capabilities: By utilizing the ORM concepts of cost-benefit analysis, the Air Force war fighters will make better, more informed battlefield decisions and will in turn help ensure victory in combat situations.

Responsibilities

While all personnel are responsible for their own use of ORM principles, different levels of each service have more specific, greater encompassing responsibilities.

AF Responsibilities.

Responsibility for the implementation of ORM practices fall on all members of the Air Force, starting from the top levels of AF Headquarters down to the individual personnel at the unit level. Table 2 summarizes the basic responsibilities for the various levels of the Air Force.

Table 2. AF ORM Responsibilities (Department of the Air Force, 2000c)

Unit Level	Responsibilities
Headquarters USAF	Advocate ORM, develop implementation and sustainment plans, appoint program focal point
Safety Center	Lead agency for integration, designate overall ORM Program Manager, guidance and oversight for all program policies
Academy, Training Command	Integrate ORM into training and education programs
MAJCOM	Develop command specific guidance
Commanders	Tailor basic ORM program guidelines to specific missions, develop implementation and sustainment plans for units
Functional Program Managers	Integrate ORM processes into unit training programs
Personnel	Apply ORM principles and practices into day-to-day activities

The AF has also formed a number of teams to help ensure the propagation of the ORM program. The AF ORM Steering Committee, co-chaired by the AF Assistant Vice Chief of Staff and the Deputy Assistant Secretary for Environment, Safety, and Occupational Health, meets annually and provides senior leadership with review and approval of ORM policy and strategy. The AF ORM Integrated Process Team is chaired by the ORM Program Manager, meets semi-annually, and develops plans needed to facilitate AF-wide implementation of ORM. An AF ORM Working Group consisting of MAJCOM ORM representatives brings MAJCOM requirements together to assist in policy development (Department of the Air Force, 2000b).

Army Responsibilities.

The Army utilizes a similar structure for the responsibility of their RM program. The headquarters level of the Army has overall responsibility for the protection of the force. The Secretary of the Army is assisted by the Assistant Secretaries of Installations and Environments and Financial Management and the Chief of Staff of the Army for top-level program guidance (Department of the Army, 1998).

The Director of Army Safety oversees direct management of the Army aviation accident prevention program. The director also runs the Army Safety Center, which is the focal point for Army aviation accident investigations, research and analysis of aviation accidents, and development of risk management options for commanders (Department of the Army, 1998).

Major Army Command (MACOM) commanders develop risk decision authority level policies. MACOM commanders including Training and Doctrine, Forces, and

Materiel commands develop specific guidance for their areas of responsibility (Department of the Army, 1998).

Commanders ensure that their units comply with all aspects of published safety guidance, including implementation of programs, training, evaluations, and establishing a written unit safety philosophy. Commanders are aided by numerous personnel assigned to various safety responsibilities. Table 3 outlines these positions.

Table 3. Army RM Responsibilities (Department of the Army, 1998)

Personnel	Responsibilities
Operations Officers	Ensure aviators are trained and briefed on mission hazards and safety procedures
Aviation Safety Officers	Develop unit safety policy, objectives, and integration procedures
Aviators	Basic element in aviation safety, must be proficient and fit and incorporate unit safety procedures
Aviation Maintenance Officers	Ensures effective maintenance program is developed and maintained
Flight Surgeon	Advises commander on all medical safety issues
Aviation Safety NCO	Advises Aviation Safety Officer, liaison with enlisted personnel

The army aviator is the key element in the aviation safety process, but all individuals are ultimately responsible for understanding safety principles and incorporating them into day-to-day activities and for advising others about unsafe actions.

Risk Management Implementation

It is useful to understand the methods and dates of ORM implementation to determine their effects on the study.

AF Implementation.

The Air Force began implementation of ORM in 1996 following the order of the Chief of Staff on 2 September 1996. The Air Force places responsibility of integrating risk management at all levels; commanders, staff, supervisors, and individuals. AFPAM 90-902 provides a brief overview of each level of responsibility, for example, individuals should 1) understand, accept, and implement risk management processes, 2) maintain a constant awareness of the changing risks associated with the operation or task, and 3) make supervisors immediately aware of any unrealistic risk reduction measures or high risk procedures (Department of the Air Force, 2000b).

The Air Force delineates the levels of risk management based on a time-criticality factor. The levels are; time-critical, deliberate, and strategic. Time critical refers to decisions that must be made at the time of execution, for example, actual mission operation or off-duty safety scenarios. Time-critical situations do not allow for the complete application of the ORM process to occur, and therefore calls for an on the spot mental or verbal review of the situation. Deliberate risk management is not time sensitive and allows for the application of the complete process. Examples of deliberate risk management can occur while planning upcoming operations. Strategic risk management is deliberate risk management augmented with more thorough identification of hazards and procedures by data analysis and research. Examples include the development of new weapon systems or tactics and training methods.

Feedback and evaluation of the ORM program is essential. By taking direct measure of behavior, conditions, attitudes, knowledge and safety statistics, a commander can ascertain how effectively his unit is incorporating the ORM principles.

Army Implementation.

According to FM100-14, the Army began to incorporate the principles of risk management in the late 1980's, where it was primarily the responsibility of the officer corps. In 1987, the Army published AR 385-10, *The Army Safety Program*, which was the Army's first formal effort at risk management (Department of the Army, 1998).

General Dennis J. Reimer authorized the release of FM 100-14 in 1998, providing the Army with a new and comprehensive risk management program. The Army clearly places responsibility for safety on all of its individuals; "Minimizing risk—eliminating unnecessary risk—is the responsibility of everyone in the chain of command (Department of the Army, 1998)." FM100-14 outlines responsibilities for differing levels of authority, from commanders and leaders to staffs and soldiers. Each level is faced with unique circumstances where the implementation of risk management is necessary and must have an ingrained understanding of the process to carry out the mission as safely as possible.

The integration of risk management into both training and operations is important and must not be treated as an afterthought. FM100-14 directs leaders and managers to account for its implementation in the beginning of the budgeting and planning process. They must also ensure constant assessment tools are in place to continually track performance (Department of the Army, 1998).

Summary

This chapter provided an overview of aviation safety and its relevance to operational risk management principles. It began with a model and a discussion of aviation mishap factors, collating the various mishap causes into four distinct mishap factors; human, environmental, material, and other. A discussion of prevention

techniques ensued, including leadership, mishap investigation, human factors programs, and technological improvements. The chapter then identified the critical terms and concepts and defined them as they pertained to the Air Force and Army ORM programs. Finally, a discussion of ORM implementation was provided, describing the differences and similarities between the Air Force and Army policies.

Through this literature review, it is evident that the Air Force has implemented ORM to instill an atmosphere of safety throughout its ranks and in particular, in the hopes that it will reduce its aviation mishaps. The next chapter will describe how various aviation mishap data was analyzed to determine whether ORM was successful or not.

III. Methodology

Chapter Overview

This chapter focuses on the methodology used to answer the investigative questions. First, a discussion of the research design is presented and threats to validity and reliability of the findings are examined. Then, the focus shifts to identify and explain the various statistical tools, tests, and procedures that were employed.

Research Design

This experiment was a quasi-experimental, time-series design. It was not a true-experiment as there was no control group available. A time-series design has a series of initial observations that take place over a period of time, interrupted with a treatment, and followed by another series of observations. The treatment being studied in this experiment is the implementation of ORM. The design is depicted diagrammatically in Figure 2.

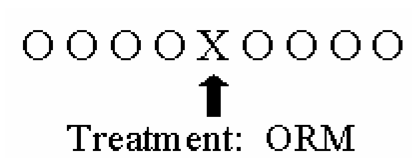


Figure 2. Research Design Diagram (Leedy and Ormrod, 2001)

Quantitative Design.

This thesis was primarily a quantitative research design, focusing on deductive analysis and adhering to the distinguishing characteristics of such designs as described by Leedy and Ormrod (2001). The methods utilized to answer the problem statement and its associated investigative questions involved studying the relationships of measured

variables and in particular, mishap rates. Its purpose was to examine the causes of mishaps, develop a model using those factors, and to test the hypothesis that ORM did not effectively reduce mishap rates.

Data Issues

Several types of data were collected from a representatively large sample of the population. The primary source of data was the Air Force Safety Center (AFSC). Historical mishap rates and summary data were obtained from the AFSC website (Air Force Safety Center, 2002). This included Class A, B, and C mishap rates and counts. AFSC database analysts provided additional mishap data, including causal counts, monthly mishap rates, and sortie numbers (Air Force Safety Center, 2003b). Similarly, Army mishap rates and summary data were obtained from the Army Safety Center website, and mishap cause counts were provided by Safety Center analysts (Army Safety Center 2002). Monthly flying hours and sorties, as well as individual mishap data were not available.

Validity and Reliability

Experiments are subjected to a number of threats to validity and reliability. This section addresses and describes a number of selected, pertinent threats and any methodologies used to counter them.

Construct Validity.

A construct is a complex, inferred concept. In this study, theory states that risk management practices affect the likelihood of aviation mishaps. The two main constructs are the management practices and the likelihood of mishaps. Construct validity, the first step in assuring a viable experiment, is a measurement of validity that "assesses the

extent to which the measure reflects the intended construct (Dooley, 2001)." Common problems with construct validity include measurement threats such as excessive random error and incorrectly measured constructs. This project intends to measure risk management's impact through statistical analysis of mishap causal data. Further threats to construct validity are the experimental threats of attrition and mortality. Since many aviation mishaps end in pilot fatalities, these threats are pertinent and may affect results.

Internal Validity.

Internal validity, defined by Dooley (2001) as the truthfulness of the claim that one variable causes another, is an essential element in any research effort. Leedy and Ormrod (2001) refer to it as the extent to which the design and data of the research allows the researcher to draw accurate conclusions about the cause and effect and other questions. If internal validity is obtained throughout an experiment, a legitimate causal linkage between the response and treatment variable is assured. Otherwise, changes in the response variable could be due to another, unexplored cause. In this research, the mishap rate is the response variable and risk management is the treatment. The primary investigative objective is to determine if the treatment causes a significant change to the response variable.

Internal validity can be threatened by time related problems, group errors, and reverse causation (Dooley, 2001). Time threats refer to rival causes other than the treatment variable that can affect the variable being measured and includes history, maturation, instrumentation, and pretest reactivity. Group threats include selection, regression to the mean, and selection-by-time threat interactions. Reverse causation is a

circumstance where the treatment variable is caused by the response variable—the opposite effect of the hypothesized relationship.

History.

History, a time threat to internal validity, is the single largest threat to this research effort. History threats occur when events unrelated to the experimental treatment cause observed reactions from the response variable (Dooley, 2001). Risk management was instituted as a means of preventing mishaps, but it is not the only effort put forth by the services to do so. As discussed in Chapter II, other programs have been studied and used to make flying safer, such as the Crew Resource Management program, mishap investigations, and leadership initiatives. These activities, which have been used for many years, are time threats to the hypothesized variable relationship. However, as Ashley (1999) noted, such programs are together to be considered responsible for trends before the implementation of ORM for the Air Force in 1996 and RM for the Army in 1987. After implementation, ORM and RM bear the weight of any cause and effect relationships that may be observed. An historical overview of such historical threats was conducted.

Conflicts.

It is possible that US involvement in military conflicts could affect flying safety. Wartime activities are accompanied by surges in operations and flying hours and puts many pilots into stressful combat situations. It would seem likely that under such situations, the likelihood of incurring increased mishaps would increase, but this concept is not supported by data.

A review of mishap rates during recent American conflicts do not show a corresponding increase in mishap rates. Table 4 illustrates mishap trends during conflicts since the Korean War in 1950 to 1953.

Table 4. Mishap Trends During Conflicts (Air Force Safety Center, 2002)

Conflict	Years	Mishap Rates	Trend
Afghanistan/Iraq	2001 to present	1.16 to 1.52	Increasing
Kosovo	1999	2.48 to 1.57	Decreasing
Gulf War	1991	1.82 to 0.82	Decreasing
Vietnam	1959 to 1975	8.29 to 2.77	Decreasing
Korea	1950 to 1953	36.48 to 24.42	Decreasing

The data from Table 4 seems to show that flying safety improves during times of conflict. Only during the current operations in Afghanistan and Iraq did the AF mishap rates increase. All other major conflicts saw improved mishap rates. Class B mishaps increased during Kosovo.

Aircraft.

Not all aircraft are created equal, and not all aircraft have the same roles in the AF. Clearly, the single-engine, high-speed F-16 with a combat role leads a much more dangerous existence than the four-engine, slower moving C-141 with a non-combat role. For this reason, it was useful to examine the different airframes within the AF fleet to determine whether aircraft mix would have any affect on mishap rates. The AF's ten aircraft with the highest Class A mishap rates over the last ten years were: U-2 (8.51), H-53 (8.49), F-117 (4.62), H-60 (3.48), F-6 (3.35), F-111 (2.84), F-15 (2.04), T-43 (1.57), E-4/E-8 (high rates, but small sample size; low significance (Air Force Safety Center, 2003a).

Not surprisingly, the mishap leaders were predominantly a mix of fighters and helicopters.   Not a single transport made the list, and only one trainer (T-43). The F-4, which began to phase out of the fleet in the late 1990's had a history of high mishap rates. It's lifetime Class A mishap rate was 4.64 (Air Force Safety Center, 2002). The F-4's removal should make for a safer mix of aircraft and reduce mishap rates overall.

More data needs to be collected and more studies need to be accomplished on the subject of aircraft mix and its affects on flying safety. It is assumed that modern airframes are better designed, have more advanced systems, and more reliable manufacturing processes. These advancements are likely to have contributed greatly to the historical reduction in the AF's and Army's mishap rates, although to what degree is unknown. One might assume that today's modern aircraft mix would contribute towards driving down mishap rates. The issue of ageing aircraft, which is a topic of study unto itself, must also be considered. Many of the AF's airframes have been in service for decades. It seems logical that as an aircraft ages, it would eventually become less reliable, and could ultimately contribute to a mishap. The small proportion of parts and manufacture related mishaps, however, does not point to this area as a serious threat.

Personnel.

Human factors contribute to the majority of Class A mishaps. We must therefore consider the historical makeup of the personnel involved in aviation mishaps. Specifically involved in an aviation mishap are pilots, maintenance personnel, and supervisors.

Pilot retention problems are well known in the AF. It seems logical that if the AF were losing pilots to the civilian sector, she would be forced to hire new ones, driving the

overall experience level and age of the pilot pool down. If this were the case, it would seem likely that mishap rates might increase, since youth and inexperience are logically linked with an increased likelihood of mishaps. Analysis of pilot data however, which is discussed in greater detail later in this chapter, shows that the pilot pool is in fact getting older and more experienced, which would lend itself to a decreased likelihood of mishaps.

The aircraft maintenance field is experiencing its own retention problems. A RAND Corporation study conducted in 2002 revealed that authorizations for enlisted aircraft maintenance personnel fell by 12.5 percent. And while fill rates of basic apprentice level crew chief maintainers (3-Levels) rose to 134 percent and supervisor crew chiefs (7-Levels) rose to 111 percent, mid-level technicians (5-Levels) fell to 75 percent. (Dahlman and others, 2002). This overall reduction, most notably in well-trained, mid-level technicians could contribute to an increase in maintenance related mishaps. However, this would be a very minor contribution, since only 4.7% of mishaps are maintenance related over the last ten years (Air Force Safety Center, 2003b)

Maturation.

Dooley defines maturation as a time threat to internal validity in which the internal processes of the experiment cause any observed changes (Dooley, 2001). In this case, it refers to the development of the pilot throughout their flying careers. Maturation is a threat to validity in this experiment due to the prevention programs utilized by the services, training, safer technology, and general experience. Since ORM was designed to reduce mishaps, one may assume that over time, the subjects individually and as a whole would achieve greater understanding of risk management principles and eventually

reduce their likelihood of being involved in a mishap. This would serve to drive down mishap rates over time.

Conversely, over time, older, experienced pilots are removed from flight status and are replaced with new, inexperienced ones, presumably resulting in a steady demographic population. The previously discussed maturation effects would be consequently nullified. Analysis of mishap demographics, however, indicates that since 1996, the sample population got older and more experienced, which would seem to contribute to a decrease in mishaps.

Mortality.

Mortality refers to the loss of test subjects due to any number of reasons, including death and voluntary removal from the sample (Dooley, 2001). Unfortunately, since many of the aviation mishaps studied in this research involve pilot fatalities, mortality is indeed a threat. It is possible that such incidents may also relate to the maturation concept. Mortality involves the removal and eventual replacement of a pilot whose attrition was most likely the result, at least in part, of human error. If an aviator were removed from the sample in this manner, it would, in effect, raise the overall level of safety for the remaining sample and could minutely lower the likelihood of future mishaps and consequently lower subsequent mishap rates. Over time, this threat could theoretically be responsible for the gradual reduction of risk. Additionally, retention problems driven by lucrative civilian flying jobs contribute to test subject attrition.

Instrumentation.

Instrumentation threats occur when there are shifts in the methods in which data is collected (Dooley, 2001). Changes in such methods are likely to adversely affect the

validity of the measured result. Minor instrumentation threats are evident in this research as the criteria for mishap classification C was modified for dollars lost slightly in 2000. The classification adjustment was minor and would not significantly change the affected rates. An additional confound was noted and studied by Ashley. Previous to 1983, the Army included Flight-related mishaps along with Flight mishaps in its rate calculations. Ashley studied the confound, concluding that the change in instrumentation was not a significant factor affecting mishap rates.

Test Reactivity.

Test Reactivity refers to a change in the subject's behavior after being exposed to an initial pretest (Dooley, 2001). It is likely that subjects would learn from any such pretest and it would adversely affect the results of the primary test. Test reactivity is not considered a threat in this research because pretests were not conducted.

Group Threats

Group threats are alternate explanations of an observed phenomenon caused by differences between studied groups rather than the treatment applied by the researcher (Dooley, 2000). Creating equivalent groups prior to experimentation alleviates these threats. In this experiment, however, we are unable to form a control group, and some threats must therefore be considered.

Two notable threats arise when a control group is not available. The first threat is that the sample does not adequately represent its parent population. In this case, however, the sample under scrutiny is the entire population of Air Force and Army aviators and is therefore a complete representation of the parent population. The second threat is that the demographics of the population may have shifted over time. It is

possible that over time, sample demographics such as age and experience may have changed. To study this possibility, an analysis of mishap demographic data before and after ORM implementation was conducted. The mean age of aircrew involved in Class A and B mishaps prior to 1996 was 30.61 years. This increased to 31.88 years for mishaps after 1996. Additionally, the mean flight hours of experience prior to 1996 was 1739.19 hours, which increased to 1894.30 hours. The average post-ORM mishap, therefore, involved slightly older, more experienced aviators. Due to the affects of maturations, older, more experienced pilots should not negatively affect mishap rates and should not have negatively skewed the results of the ORM program, unless of course, such pilots adopt a more cavalier approach towards safety.

Since these two threats do not appear to directly affect the population sample, group threats are not considered a threat to the validity of the research.

Selection.

It is essential that the selection of the experimental groups be accomplished fairly and appropriately. It is possible that selected groups may differ in certain regards prior to the experiment and this may pose a threat to internal validity. Selection is a group internal validity threat defined by Dooley as "differences observed between groups at the end of the study existed prior to the intervention because of the way members were sorted into groups." (Dooley, 2001) Since control over groups was not possible in this research, the entire group is being studied. Selection, therefore, is not considered a threat.

Selection-By-Time-Interactions.

Selection-By-Time-Interactions refer to situations in which subjects with different chances of observing time related changes, such as maturation or history, are located within different groups (Dooley, 2001). All Air Force and Army pilots and their mishap rates are being studied conjunctively in this research and are presumably exposed to very similar time related changes. The selection-by-time-interaction threat is therefore considered minimal.

Regression Towards the Mean.

Regression Towards the Mean is a group threat in which extremely high and low responses are grouped together and retested, gravitating towards the mean observation and subsequently resulting in less extreme results (Dooley, 2001). In this case, statistical regression analysis is used to study data, and extreme mishap rates and data outliers are removed when appropriate. For this reason, the regression towards the mean threat is considered minimal.

Reverse Causation

A research design that measures a number of variables concurrently runs the risk of reverse causation, in which the cause and effect relationship of the variables is not properly determined and temporal precedence of the variables is not understood (Dooley, 2001). It is possible to determine correlations between such variables, but if a response variable were not set before a treatment variable was administered, reverse causation would be a threat. In this case, ORM practices were implemented long after the rates of aviation mishaps were being monitored, and indeed, rates were already going down prior to ORM implementation. Therefore it is not likely that ORM was implemented in

response to a change (up or down) in accident rates. Additionally, the statistical methodology employed explicitly uses the temporal precedence of ORM through the use of piecewise linear regression. Consequently, reverse causation is a mild threat to this research design.

Statistical Inference Validity

Statistical inference validity is tested by inferential statistics, and is obtained when the likelihood that the findings of the experiment are due to mere chance can confidently be dismissed (Dooley, 2001). It is possible that the results of an experiment are due to errors in data sampling, such as improper population sampling or a small data sample. In this case, flight mishap statistics are the critical element of this research, and its validity as proper measurement data is clear. Sample sizes are quite substantial when broken down into quarterly data. Statistical inference validity is not considered a threat to this research.

A possible source of error is that this research studies only failed sortie data (mishaps). A more useful data source would be a database of both successful and failed sorties and their associated statistics. It would be useful to compare the two populations and it would eliminate the threat of the successful sortie population being different than the failed population.

Additionally, this research uses a combination of parametric and non-parametric data. The methodologies used to analyze such data vary. Where the delineation between parametric and non-parametric is not clear, both types of tests are used.

Time series data is also a possible source of validity threat. Conversion of the time series data into a percentage period index and exponentially smoothed data alleviates the threats.

External Validity

Whereas internal validity pertains to the relationships within an experimental study, external validity refers to the generalizability of the research's findings to external populations, places, or times, and always involves the interaction of the treatment with some other factor (Dooley, 2001). Ashley's determination that ORM would not reduce the Air Force's mishap rate is an external extension of his findings of the Army's program (Ashley, 1999). Findings from this study would confirm the external validity of those findings to other populations; in this case, Air Force pilots. A source that could be used to test the external validity of both Ashley's conclusions and this research is the U.S. Navy mishap rates and RM program. Findings from this study would not be generalizable to non-military aviation, however. There are considerable differences between military flying and commercial or general aviation. The inherent external validity threat in this case is disregarded, as this thesis is only concerned with findings pertinent to military aviation.

A summary table of the threats to research validity is shown in Table 5.

Table 5. Threats to Validity

THREAT	LEVEL	DESCRIPTION/WORKAROUND
History	Medium	Many unknown factors possibly involved/ Perform tests around suspected factors
Maturation	Medium	Deviation towards safety after implementation/ None
Mortality	Medium	Observations are often fatal/ Examined demographics
Instrumentation	Low	Insignificant Class C data shift
Selection	Low	Entire population
Regression	Low	Outliers are not retested
Testing	Low	No pretest to react to
Reverse Causation	Low	Decrease in rates did not cause ORM
Statistical Inference Validity	Low	Data is non-parametric, small sample size, time series/ Use numerous tests, smooth times series data
External Validity	Low	AF pilots are not the same as GA, commercial pilots/ NA; only care about military pilots

Investigative Questions

The following section discusses the methodology of each of the five investigative questions.

IQ.1: What are the factors involved in an aviation mishap?

This investigative question is answered in Chapter 2, Literature Review.

IQ.2: What is ORM and how is it implemented?

This investigative question is answered in Chapter 2, Literature Review.

IQ.3: Have mishaps rates changed significantly since ORM was implemented?

This statistical analysis sought to detect significant differences in the mishap rates before and after the implementation of RM programs. The Air Force began its ORM program in 1996, so mishap rates from FY 1983 to 1996 were compared to those of FY 1997 to 2002. Ashley's investigation determined that the Army showed no significant

improvement after 1987 when their similar program was implemented (Ashley, 1999). A comparison using updated Army mishap rates from 1973 to 2002 was accomplished to validate his results. To determine any significant changes, a number of comparison of means tests and comparison of variance tests were conducted.

Comparison of Means.

The methodology of this phase is based on comparisons of population means from small sample sizes, due to the relatively small number of data points (Anderson and others, 1999). Three assumptions must be met to perform the comparison tests (Devore, 2000). The first assumption is that both samples must be selected from populations with normal probability distributions. The second is that the samples are independent and randomly selected. The third is that the samples must be taken from populations with equal variances.

The first assumption was satisfied through an analysis of the residuals. Residuals, as defined by Anderson and others, are the difference between the observed value of the mishap rate and the value predicted using the estimated regression equation (Anderson and others, 2000). To determine residuals, a linear regression was performed using the mishap rate as the dependent variable and fiscal year as the independent variable. Results are shown in Appendices C, D, E, and F. An analysis of the data residuals using the Kolmogorov-Smirnov (K-S) goodness of fit test verifies this requirement. The K-S test is used to test the hypothesis that a sample comes from a particular distribution (normal in this case). The value of the K-S Z statistic is based on the largest absolute difference between the residual and the theoretical cumulative normal distributions.

The second assumption is that the samples are independent and randomly selected from their populations. To truly satisfy this assumption, it would be necessary to have access to comprehensive data from all flights—both successful sorties and failed sorties (mishaps). Unfortunately, comprehensive data of this nature is not available, and we are left with only the failed sortie data. However, this assumption was satisfied because the sample is composed of all available data points of the failed sorties for the population being studied.

The third assumption is that the samples must be taken from populations with equal variances. The mishap rates being studied are time series data, however, so a test of variances is not appropriate, and the methodology for comparing the means must be reevaluated. To that end, the data was transformed using a percentage period index method and exponential smoothing, both of which are discussed later in the chapter. Once transformed, direct comparison of means is applicable.

The mishap rates are a chronological sequence of observations on a single variable and can be therefore defined as time series data (Bowerman and O'Connell, 1999). Time series can be either stationary or non-stationary. A time series is stationary if it fluctuates around a constant mean. The studied mishap rates, however, do not fluctuate around a constant mean and are therefore considered non-stationary. Non-stationary time series must be transformed into stationary time series before comparisons of means may be performed.

Percentage Period Index Transformation.

To transform the data into a stationary time series, the percentage period index (PPI) procedure used by Ashley (Ashley, 1999) and described by Makridakis was employed (Makridakis, 1983). The PPI is a period-to-period percentage change measurement that enables the computation of testable means by converting the non-stationary means into stationary PPI means. Testing the differences of the PPI means will determine whether there was a significant difference after RM implementation.

The PPI transformation begins with setting the value of the first year's mishap rate to a constant, *C*, in order to create an order of magnitude for the index. PPIs for subsequent years are then calculated by determining the ratio of the current mishap rate to the previous year's mishap rate and then multiplying the result by the selected constant. The PPI formula with a selected constant, *C*, of 10 are calculated as follows:

$$PPI = [(Rate_{i+1}) / (Rate_i)] \times C \qquad from\ i = 1\ to\ n \qquad (1)$$

Resulting tables of PPI values are included in Appendices G and H.

Once the mishap rates were transformed, comparisons of means tests were conducted.

Time Series Data Transformation: Exponential Smoothing.

A second transformation, known as exponential smoothing with trend adjustment, was used to adjust the time series data. This algorithm works by smoothing out blips in the data while adjusting for a trend over time. Smoothing the data set allows analysis that is less susceptible to the influence of extreme values.

This methodology creates a smoothed value (S_t) of the actual observation (A_t) by adjusting for trends (T_t). Two smoothing constants, α and β, are applied in the formulation and can fall between 0.1 and 0.5. The median of 0.3 for both values was chosen for this study.

The formula of the smoothed trend is:

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1} \qquad (2)$$

The formula of the smoothed value is:

$$S_t = \alpha(A_t) + (1 - \alpha)(S_{t-1} - T_{t-1}) \qquad (3)$$

The calculated smoothed values replace the original rates and are then analyzed using comparison of means tests explained hereafter. The exponential smoothing values are shown in Appendices O and P.

Test Descriptions.

To determine whether there was a statistically significant difference in means before and after implementation, a number of tests were conducted using the SPSS® 8.0 statistics package. To illustrate the differences between the raw mishap rates, trend adjusted PPI rates, and moving average adjusted rates, tests were conducted on all three sets of data. A simple examination of means of the actual rates showed decreases in 3 of the 4 data categories studied, as shown in Table 6.

Table 6. Mishap Rate Simple Means Comparisons

	Pre-ORM	Post-ORM	Trend
AF Class A	1.543	1.294	Decrease
AF Class B	0.549	1.99	Increase
Army Class A	2.873	1.639	Decrease
Army Class B-C	13.481	7.306	Decrease

A series of charts showing these rates, adjusted PPI rates, and moving average rates over the examined time period and results from the tests will be shown in Chapter 4.

Parametric Tests.

The first two tests, ANOVA and T-Tests are parametric tests. They rely on the assumption that the samples come from populations that follow a normal distribution and are from a continuous interval or ratio scale (Devore, 2000). While it is not appropriate to test the normality of the actual mishap rates, analysis of the data residuals showed that they were from approximately normal distributions (Appendices C-F). Additionally, mishap rates are continuous interval scalar values. Therefore, parametric tests may be appropriate.

ANOVA.

ANOVA tests compare means of different samples through analysis of variance. The test statistic for ANOVA tests is the F-statistic. The F-statistic is computed by dividing the mean square due to treatments by the mean square due to error. The F-statistic is compared to a critical F-value to yield a p-value. Large F statistics yield small p-values, which must be less than the test's alpha value to reject the null hypothesis at the desired confidence level (Devore, 2000).

T-Test.

The T-test is used to determine statistically significant differences in the means of two groups. The test calculates a t-value by dividing the difference in means between the two groups by its standard error. Large t-values result in small p-values (Devore, 2000).

Non-Parametric Tests.

The remaining two tests, the Mann Whitney Test and the Wilcoxon Sign-Rank Test, are non-parametric, which alleviates the requirement for sample normality and continuous interval values (Devore, 2000). Due to the difficulties of defining time series mishap rate data, these non-parametric tests were used as an additional, independent check on the validity of inferences drawn from the parametric tests.

Mann Whitney Test.

The Mann Whitney test is used to determine statistically significant differences in the means of two groups. This test is used for non-parametric populations, useful when standard assumptions about population distributions are not applicable (Devore, 2000). The test statistic for the Mann Whitney test is the U statistic, with large values yielding small p-values.

Wilcoxon Sign-Rank Test.

The Wilcoxon Sign-Rank Test is used to determine statistically significant differences in the means of two groups. It is used for non-parametric populations (Devore, 2000).

Comparison of Variances.

The comparison of variances, like the comparison of means, is problematic when using time series data. To compare variances of the mishap rates appropriately, an

analysis of the residuals of the mishap rates when regressed against the fiscal year may be conducted. Changes in the variances of the samples from before and after implementation may indicate that a process change had occurred. A simple glance of the mishap rate charts in Chapter 4 (Figure 23) shows a considerable amount of variance for the Army data, but is inconclusive when looking at the AF data. The AF Class A data seems to consistently vary from year to year, while the Class B data fluctuates considerably. Statistical tests of the residuals will yield more definitive answers.

When comparing variances of two samples, inferences may be made from the ratio of the variances. The null hypothesis is rejected when the ratio is compared to an F-value based on the size of the samples, yielding a small enough p-value (Anderson and others, 1999). The F-statistic, which is the ratio, is computed by placing the larger variance as the numerator and the smaller variance as the denominator. The critical F-value to which the F-statistic is compared is determined based on the degrees of freedom of the sample. When the variances are statistically the same, the null hypothesis is not rejected and we may not therefore conclude that any process change has occurred since implementation of ORM. The hypotheses were:

Ho: The residual variances are equal.

Ha: The residual variances are not equal.

This is a two-tailed test, so with an alpha value set at 0.05, the null is rejected with a p-value of 0.025 or smaller.

IQ.4: Are any differences caused by ORM?

To determine whether any rate changes were caused by the implementation of ORM, a statistical technique utilized in Ashley's thesis (Ashley, 1996) known as

discontinuous piecewise linear regression was performed. Discontinuous piecewise linear regression determines whether a slope or intercept change is present at a selected point in time (Neter and others, 1996).

A two variable model with a breakpoint at C is described as:

$$E(MR) = \beta_0 + \beta_1 * X_1 + \beta_2 * (X_1 - C) * X_2 + \beta_3 * X_2 \qquad (4)$$

where β_0 is the Y-axis intercept, β_1 is the slope of the line for the period prior to the treatment at breakpoint C, $\beta_1 + \beta_2$ is the slope of the line after C, and β_3 is the jump in the intercept at C. Figure 3 shows the concept.

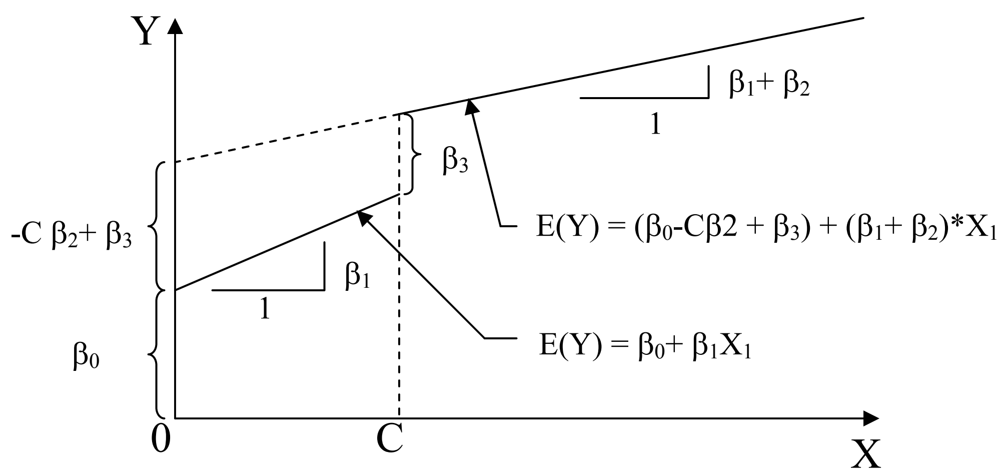


Figure 3. Discontinuous Piecewise Linear Regression Response Function (Neter and Others, 1996)

If no significant change in the slope of the regression were to occur at point C, then the two lines would have the same slope. In this case, one would expect the value of β_2 to be zero and for both lines to have a slope of β_1. If no significant shift at the intercept at point C were to occur, one would expect the value of β_3 to be zero.

With the successful implementation of the ORM treatment, one would expect to see significant changes while using these statistical procedures. An effective treatment

would yield a decreasing shift in slope and/or a decrease at the intercept at C. A shift at the intercept without a change in slope, or, conversely, a change in slope without a shift at the intercept could identify whether the treatment forced a process change (Campbell, 1963). As the AF implemented ORM in 1996, one would expect to see a downward shift at *C* or a decreasingly negative slope of the regression line after 1996.

The model consists of two variables: fiscal year (*FY*) and operational risk management (*RM*). Years prior to 1996 had an *RM* value of 0 and years after 1996 had an *RM* value of 1. The breakpoint, *C*, is 1996. The full model is:

$$E(MR) = \beta_0 + \beta_1 * FY + \beta_2 * (FY - 96) * RM + \beta_3 * RM \qquad (5)$$

where β_0 is the Y-axis intercept, β_1 is the slope of the regression line for the period prior to 1996, and β_3 is the shift in the intercept at *C*, between 1996 and 1997.

Hypotheses for the analysis were as follows:

Ho: $\beta_1 = \beta_2 = \beta_3 = 0$

Ha: The β values are $\neq 0$.

The value of the β_1 and β_3 terms are determined directly from their p-values resulting from the overall F-tests of the full model. A partial F-test must be conducted on the reduced model to determine the value of β_2. The partial F-test had the following hypotheses:

Ho: $\beta_2 = \beta_3 = 0$

Ha: $\beta_2 = 0$ or $\beta_3 = 0$, but not both

To determine if the slopes of the pre- and post-ORM regression lines are significantly different from each other, results of the partial F-test are analyzed. If the value of β_2 is

zero, then the slope of the second line will not be significantly different from the slope of the first. The resulting hypothesis was:

Ho: $\beta_2 = 0$

Ha: $\beta_2 \neq 0$

These tests and hypotheses were applied to AF Class A and B rates as well as Army Class A and B/C rates. The breakpoint, *C*, for Army data was 1987, the year RM was implemented in the Army. All tests were conducted using an alpha level of significance equal to 0.05.

IQ.5: Has the proportion of human factor related mishaps decreased since implementation?

As ORM was intended to instill an atmosphere of safety in all AF personnel, one would expect to see a reduction in the proportion of human factors, and particularly those directly affected by ORM. In this way, the experimental design would protect our results from the effects of non-ORM factor changes. To study this expectation, mishap causal count data was analyzed using the chi-square goodness of fit test for Class A and B data for both the AF and Army human factors mishaps.

Chi-Square Goodness of Fit Test.

The chi-square goodness of fit test is an upper-tailed, non-parametric test used to identify differences in observed and expected population behavior (Devore, 2000). Each category (k) being observed is assigned an expected proportion. In this case, only human factors cause categories, such as accepted risk, discipline, and emotional states were included. The test compares the proportion of actual observed instances of such causes

after implementation to a proportion based on historical averages prior to implementation.

The hypotheses are as follows:

> Ho: The population follows a multinomial probability distribution with specified probabilities for each of k categories.
>
> Ha: The population does not follow a multinomial probability distribution with specified probabilities for each of k categories.

The test statistic is the chi-square, or χ^2, and incorporates the observed frequencies (*f*) and expected frequencies (*e*) of each of k categories. The test uses *k*-1 degrees of freedom and a level of significance of 0.05. The χ^2 term is shown as:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(f_i - e_i\right)}{e_i} \qquad (6)$$

If the test statistic is shown to be less than the critical value given a level of significance of 0.05 and *k*-1 degrees of freedom, we accept the null hypothesis that the expected proportions are followed. The results of this test may provide insight into the efficacy of ORM implementation by revealing any changes in the proportion of human factors related mishaps.

Summary

This chapter explained the methodology used to answer the research question. It began by describing the research design as a quasi-experimental time-series experiment. A description of the various threats to validity was presented. Finally, the methodology

utilized to answer the investigative questions was then described. Analysis and results of the investigative methodologies are presented in the next chapter.

IV. Analysis and Results

Chapter Overview

The purpose of this chapter is to answer the overall research question by answering the five investigative questions posed in Chapter 1. For each investigative question the problem is restated, relevant data is described, and answers are presented according to the methodology described in Chapter 3.

The analysis of investigative questions 3 and 4 ultimately allows us to identify differences in the mishap rates contemporaneous with RM and ORM implementation. Investigative question 5 would discern whether the changes were also contemporaneous with changes in human factors causes. The results of the questions would provide strong circumstantial evidence that ORM and RM did or did not cause reductions in mishap rates and that it may be associated with any decreases or increases.

IQ.1: What factors are involved in an aviation mishap?

Aviation mishaps are caused by an endless list of causes such as human error, weather, bird strikes, faulty parts, etc. All such causes can be classified into one of four primary mishap causal factors: human factors, environmental, material failure, or other. These four factors, either alone or in conjunction with each other, cause aviation mishaps.

IQ.2: What is ORM and how is it implemented?

ORM is a system implemented by the Air Force in an effort to increase safety. It was designed as a decision-making process that identifies risk, evaluates courses of action, and determine the most beneficial course of action for any possible situation, on- or off-duty. It was implemented in Sep 96 and was fully integrated through AF-wide

computer training by Oct 98. Its implementation relies on commander leadership and individual adherence to its fundamental principles.

IQ.3: Have mishap rates changed significantly since the implementation of ORM practices?

Data.

The data set being used to conduct the AF comparison of means tests are Class A and Class B mishap rates from 1983 to 2002 collected from the Air Force Safety Center online database. PPI rates and moving average rates calculated from the true rates are also analyzed in the tests. The Army tests use Class A and Class B-C mishap, PPI, and exponential smoothing rates from 1973 to 2002, initially collected from the Army Safety Center online database. The Class B-C mishap rate is a combination of Class B and Class C mishaps, as provided by the Safety Center (Army Safety Center, 2002). SPSS® and Excel® were used to run the four tests. The outputs from the tests can be found in Appendices K-P.

AF Data Charts.

The following series of charts illustrates the three sets of AF mishap data: mishap rates, PPI rates, and exponential smoothing rates for Air Force Class A and B mishaps. The first chart (Figure 4) shows basic mishap rates as gleaned from the AFSC website data. Embedded trend lines indicate a slight but steady decrease in Class A rates. Class B rates were holding steady under 1.00 mishap per 100,000 flying hours until a dramatic spike occurred in 1999 and beyond.

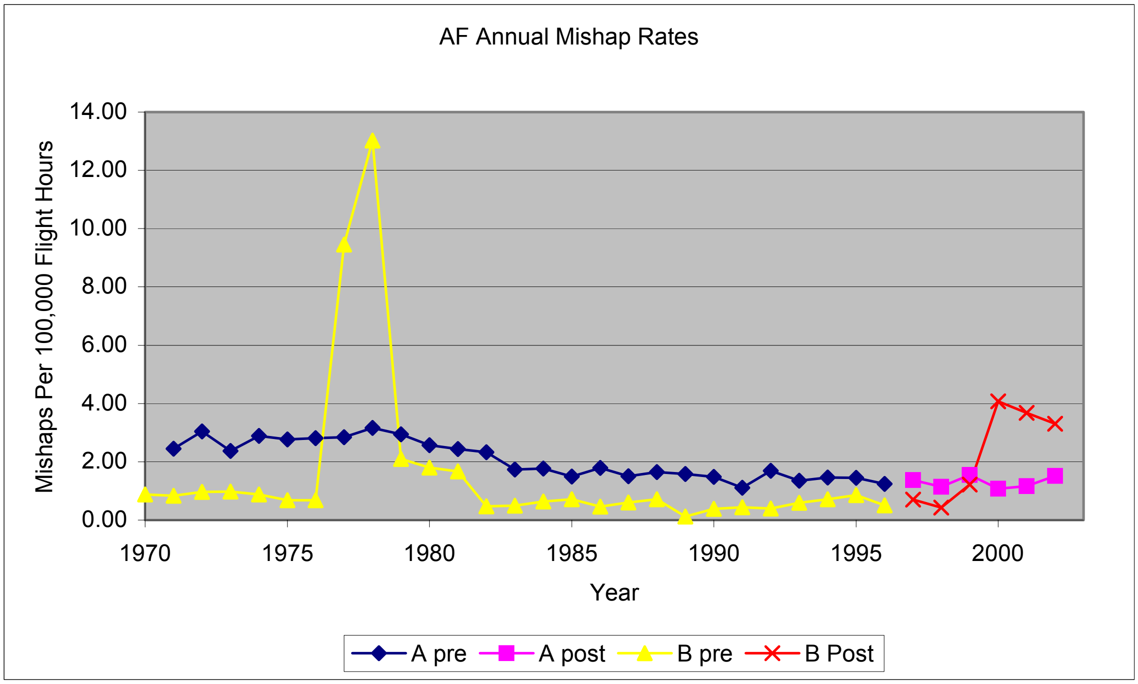


Figure 4. AF Mishap Rates

This second chart (Figure 5) illustrates the transformation of the basic rates into the PPI. As the non-stationary time-series mishap rates are anchored around a constant of 10, the once declining or steady trend lines begin to incline slightly. Pre-ORM PPI values, as indicated by their trend lines, are almost steady, with only a slight increase. Post-ORM values continue those trends with no visible change.

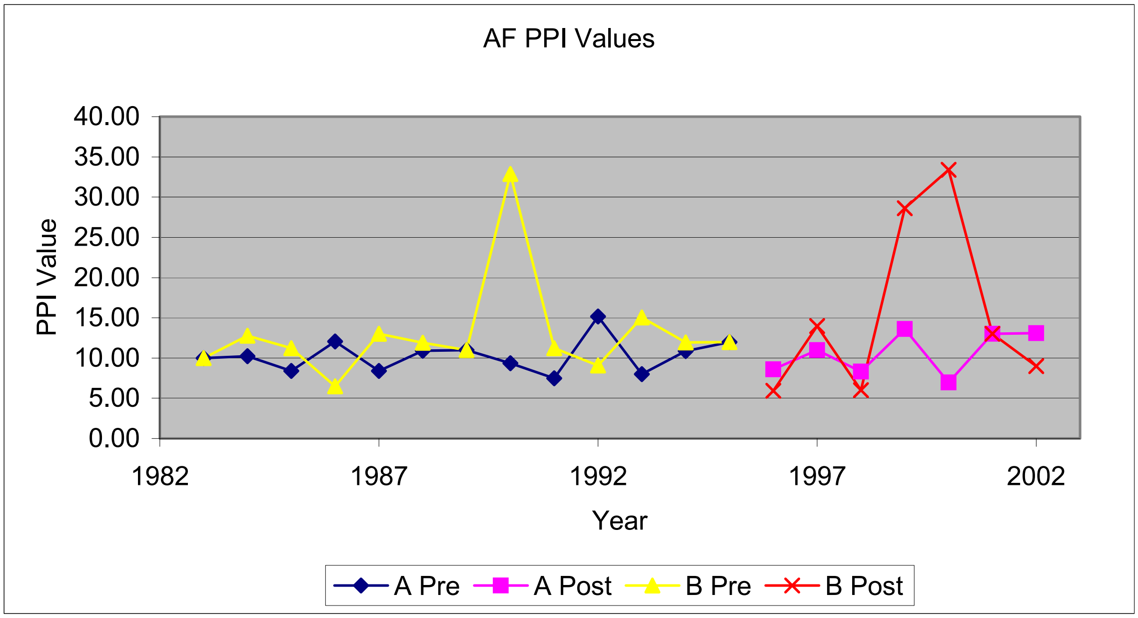


Figure 5. AF PPI Values

The third chart (Figure 6) shows the basic mishap rates transformed using exponential smoothing. Trend lines for these values indicate that the mishap rate for Class A was declining, but leveled off over the post-ORM years. Class B exponentially smoothed rates show a decrease until the start of the 1990's, when rates began to increase. A comparison of the pre- and post-ORM years for Class B indicates an increase since ORM was introduced.

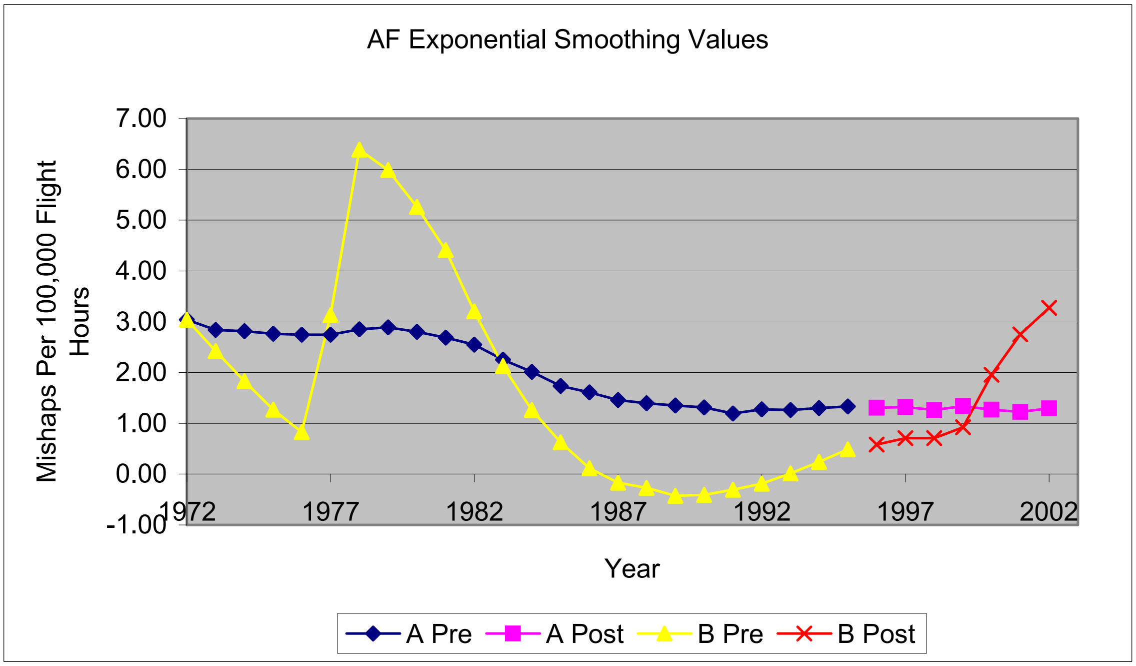


Figure 6. AF Exponential Smoothing Rates

Army Data Charts.

The next three charts display Army mishap data from 1973 to present. The charts show basic mishap rates, PPI rates, and exponentially smoothed rates for Class A and Class B-C mishaps. The first chart (Figure 7) illustrates the overall declining trends for both Class A and B-C basic mishap rates. A rudimentary glance at the chart indicates that class B-C rates seemed to have increased after RM was implemented in 1987.

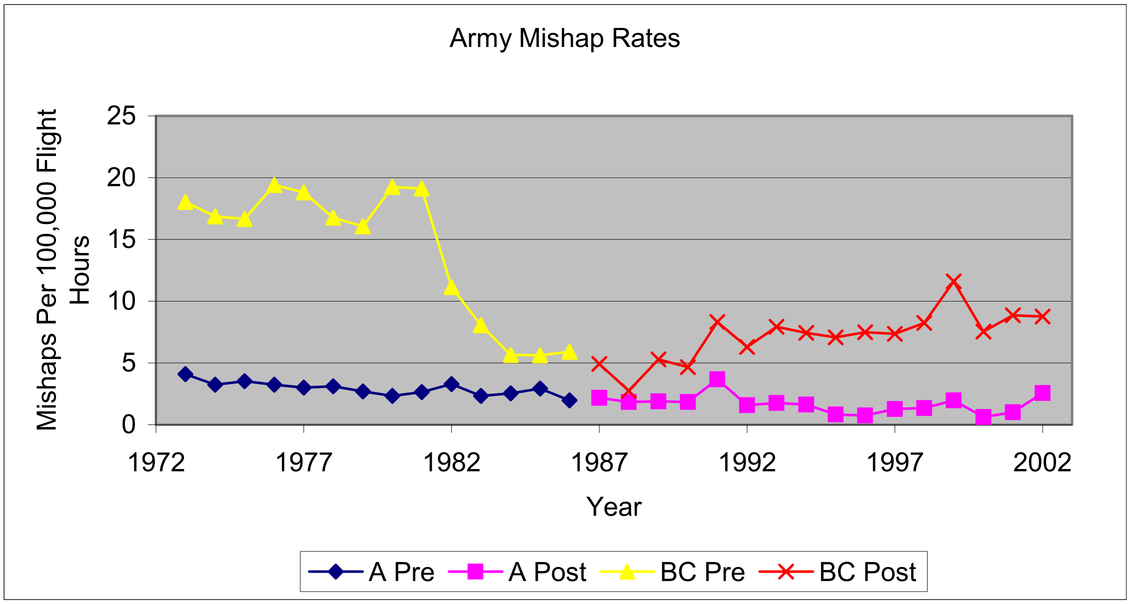


Figure 7. Army Mishap Rates

The second chart (Figure 8) shows the data after being transformed using the PPI procedure. Pre-RM values no longer show any discernable decrease, and the Class A trend actually increased after RM implementation.

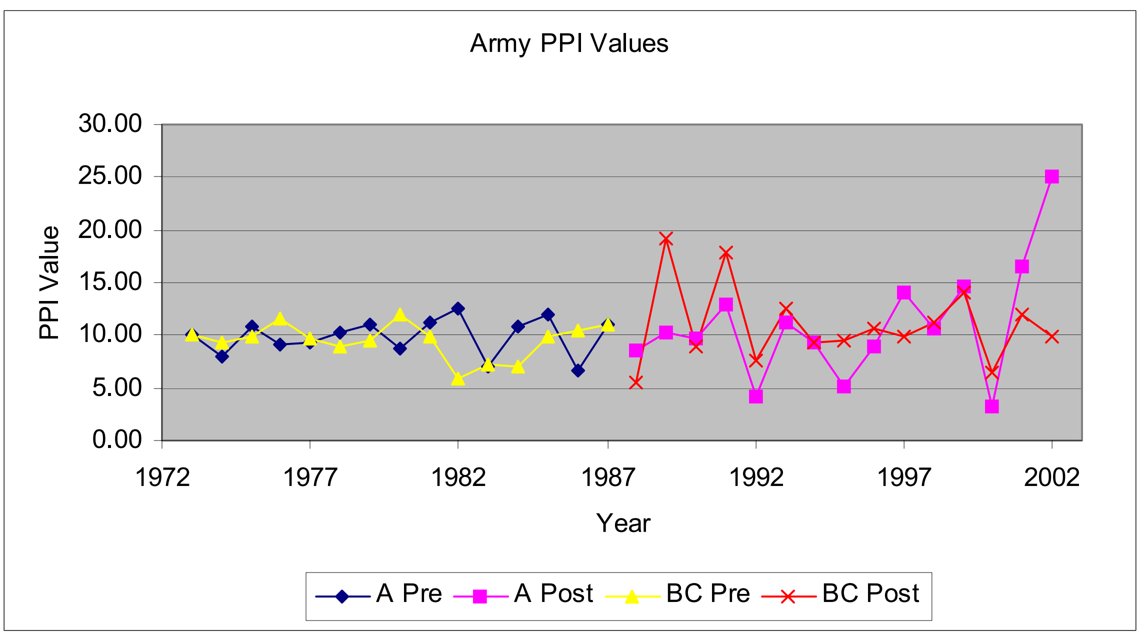


Figure 8. Army PPI Values

The third chart (Figure 9) shows the Army's mishaps rate after the application of exponential smoothing. Trends continue to follow the same pattern as the basic mishap rates. The most notable trend is the B-C rate increasing in the post-RM years.

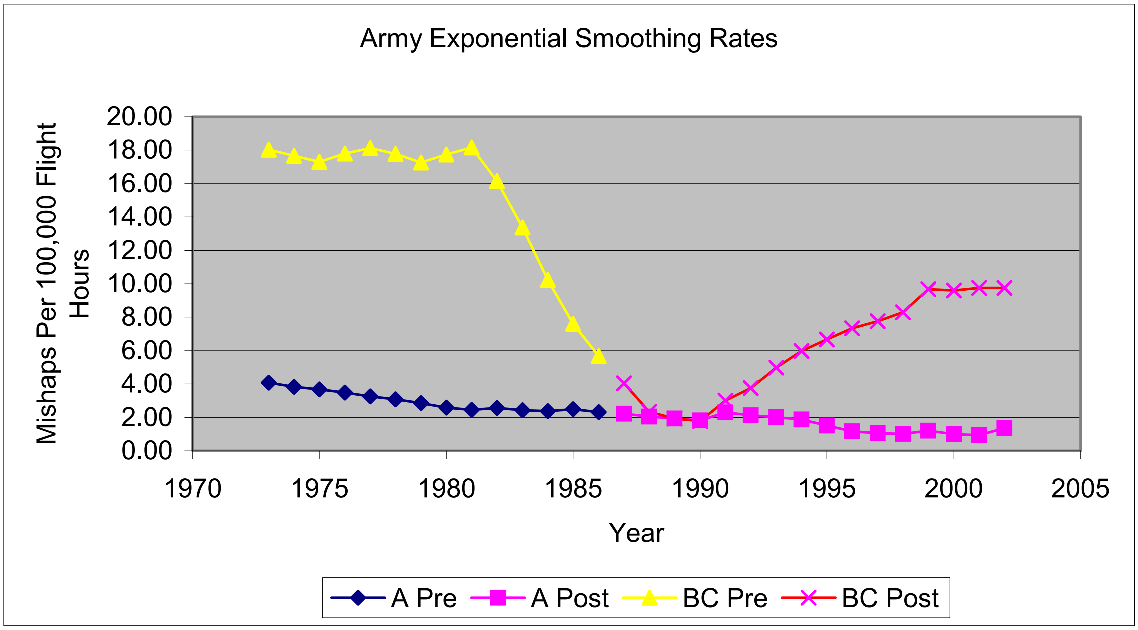


Figure 9. Army Exponential Smoothing

Results.

To determine whether the implementation of ORM had any effect on mishap rates, a series of comparison of means tests were conducted on a variety of data types. The tests analyzed whether the means of the mishap rates before ORM implementation differed from mishap rates after ORM implementation. Three data rates were analyzed; mishap rates, PPI values, and exponentially smoothed rates. Two classes were analyzed; Class A and B for the Air Force and Class A and B-C for the Army. The results are presented in the following section.

AF Comparison of Means Tests.

The results of the four tests for the AF mishap rates are shown in Table 7. Parametric tests indicate that the pre- and post-ORM years have unequal means, while the non-parametric tests, which are less sensitive and more conservative, yield somewhat different results.

Table 7. AF Mishap Rate Comparison of Means

	AF Class A		AF Class B	
	P	**Reject?**	**P**	**Reject?**
ANOVA	0.012	Yes	0.005	Yes
T-Test	0.015	Yes	0.057	No
Mann-Whitney	0.036	No	0.043	No
Wilcoxon	0.037	No	0.046	No

The results of the four tests for the AF PPI values are shown in Table 8. All tests results indicate that mean PPI values did not change after ORM.

Table 8. AF PPI Values Comparison of Means

	AF Class A		AF Class B	
	P	**Reject?**	**P**	**Reject?**
ANOVA	0.742	No	0.486	No
T-Test	0.764	No	0.561	No
Mann-Whitney	0.663	No	0.905	No
Wilcoxon	0.699	No	0.938	No

The results of the four tests for the AF exponentially smoothed rates are shown in Table 9. Tests on Class A rates indicate that the sample means did not change. Class B tests showed that means were equal.

Table 9. AF Exponential Smoothing Comparison of Means

	AF Class A		AF Class B	
	P	**Reject?**	**P**	**Reject?**
ANOVA	0.016	Yes	0.893	No
T-Test	0	Yes	0.848	No
Mann-Whitney	0.008	Yes	0.325	No
Wilcoxon	0.006	Yes	0.347	No

The tests conducted on the raw mishap rates show a significant statistical difference between Class A rates after implementation of ORM when using parametric tests, but not when using the non-parametric tests. The results indicate a possible change since implementation, and as the post-ORM mean is lower, it suggests that ORM did have its desired effect on the rates. Class B rates do not clearly show differences, although the P-values are very close to the rejection region. Due to the difficulties with the comparison of time-series data rates, the PPI tests were then conducted to yield more information.

Once trends are smoothed out using the PPI procedure, the statistical tests show no significant differences in the PPI means before and after ORM implementation. All four tests yielded p-values greater than the test level of significance of 0.05. Therefore, the tests do not reject the null hypothesis that the means are equal, and we cannot say that ORM implementation has reduced the rate of mishaps within the Air Force.

Tests conducted on the exponentially smoothed mishap rates contradict the previous findings. Class A tests unanimously rejected the null, indicating that the pre- and post-ORM means were not equal, and that a significant rate change had occurred, again suggesting a desired ORM effect. Class B tests followed the previous PPI tests by showing that the means were equal and statistically unchanged.

Overall, the Class A tests yielded contradictory results. While several of the tests showed a decreasing mean, the most reliable set of data, the PPI-transformed data, did not show a significant change. Clearly, a more expansive investigation of the rates is necessary.

Only one of the twelve tests indicated a change in the means of Class B data--the ANOVA conducted on the annual mishap data. These results do not indicate a change of means, suggesting that the implementation did not affect mishap rates. However, examination of the Figures 4 and 6 clearly indicate Class B data has taken a dramatic upswing within the last decade or so. Another glaring problem with these results is the considerable spike that happened in the late 70's, which would most likely skew the tests. Tests were therefore rerun on the Class B data with the abnormal years removed to compare results. This time the test rejected the null, indicating the means were not equal, and that rates had significantly increased since implementation. Since none of the results indicated that ORM was having its desired effect, more analysis using more sophisticated techniques was clearly needed.

Army Comparison of Means Tests.

The results of the four tests for the Army mishap rates are shown in Table 10. Results from these tests indicate that the pre- and post-RM means were not equal.

Table 10. Army Mishap Rates Comparison of Means

	Army Class A		Army Class B-C	
	P	Reject?	P	Reject?
ANOVA	0	Yes	0.001	Yes
T-Test	0	Yes	0.001	Yes
Mann-Whitney	0	Yes	0.017	Yes
Wilcoxon	0	Yes	0.016	Yes

The results of the four tests for the Army PPI values are shown in Table 11. Tests on the PPI values unanimously indicate that the rates from pre- and post-RM had not significantly changed.

Table 11. Army PPI Values Comparison of Means

	Army Class A		Army Class B-C	
	P	Reject?	P	Reject?
ANOVA	0.486	No	0.18	No
T-Test	0.49	No	0.185	No
Mann-Whitney	0.678	No	0.431	No
Wilcoxon	0.683	No	0.436	No

The results of the four tests for the Army exponential smoothing rates are shown in Table 12. The tests results unanimously indicate different means for pre- and post-RM rates.

Table 12. Army Exponential Smoothing Comparison of Means

	Army Class A		Army Class B-C	
	P	Reject?	P	Reject?
ANOVA	0	Yes	0	Yes
T-Test	0	Yes	0	Yes
Mann-Whitney	0	Yes	0	Yes
Wilcoxon	0	Yes	0	Yes

The tests conducted on the Army's raw mishap rates for both A and B-C classes show a significant difference between pre- and post-implementation of RM. Analysis of the PPI rates, however, yields different results. All four tests fail to reject the null hypothesis that the means are equal, indicating that the Army did not see a reduction in mishap rates after RM implementation. Conversely, the exponential rates tests yielded the opposite answer. All of the tests for both Class A and B-C yielded significances well within the rejection region, strongly indicating unequal means.

Comparison of Variances Tests.

The results of the four comparison of variance tests are summarized in Table 13 and the analysis can be found in Appendices Q and R. If the resultant F-statistic is

greater than the F-critical value, we may conclude that the variances are equal and no process change is likely to have occurred.

Table 13. Comparison of Variance Results

	Class	F-stat	F-crit	Reject Null?
AF	A	1.703	2.92	No
	B	7.9	2.92	Yes
Army	A	3.936	2.53	Yes
	B-C	0.765	2.48	No

The analysis of the residuals yielded varying results. AF Class A residual variance were equal and did not show a process change from pre-ORM to post-ORM, while Class B residuals did. Conversely, Class A residual variance for the Army were not equal, indicating a possible change, while the Class B-C residual variance were statistically equal.

Summary.

This investigative question sought to determine whether implementation of ORM had any significant effect on aviation mishap rates. Due to the problematic nature of time-series data, both parametric and non-parametric tests were used to compare the means of pre-ORM implementation mishap rates and post-implementation mishap rates. The results were not conclusive. While several tests conducted on mishap rates and exponentially smoothed rates showed a difference between pre- and post-ORM implementation, analysis of the PPI values clearly indicated that there was no difference. Based on these results, we cannot clearly state whether mishap rates changed or remained the same after ORM was implemented. However, since the PPI values remove the problems associated with time series data means comparisons, the PPI results are the most reliable.

The tests conducted on the Army data yielded similarly conflicting results. Tests on mishap rates and exponentially smoothed rates showed clearly that the rates had changed after implementation of their RM program. However, the same tests conducted on transformed PPI values said just the opposite; that the rates had not changed.

Analysis of the variance of the residuals of mishap rates regressed against fiscal year yielded varying results, indicating a possible process change within the data. An ORM induced change may not be discounted at this point and further analysis is required.

IQ.4: Are any changes caused by ORM?

To determine whether rate changes were due to the implementation of ORM, discontinuous piecewise linear regression was used on a number of data sets from both the Air Force and Army. This statistical technique measures changes in slope and linear shifts around a breakpoint, the date of ORM implementation.

Data.

The tests were run on several data sets, as outlined in Table 14. The AFSC provided monthly flying hours and sorties flown, enabling the development of quarterly mishap and sortie rates for additional analysis. The quarterly mishap and sortie rates were calculated as the number of mishaps per 100,000 flying hours or 100,000 sorties flown. Army quarterly data was unavailable.

Table 14. Regression Data Sets

	Annual Rates	Quarterly Mishap Rates	Quarterly Sortie Rates
Air Force	A, B	A, B	A, B
Army	A, B-C	NA	NA

Air Force Results.

AF Class A Annual Rates.

The AF Class A annual mishap rates are illustrated in Figure 10. The time periods were 1970 to 1996 for the pre-ORM years and 1997 to 2002 for the post-ORM years. The chart shows the two periods with the breakpoint, C, at 1996 along with associated regression lines.

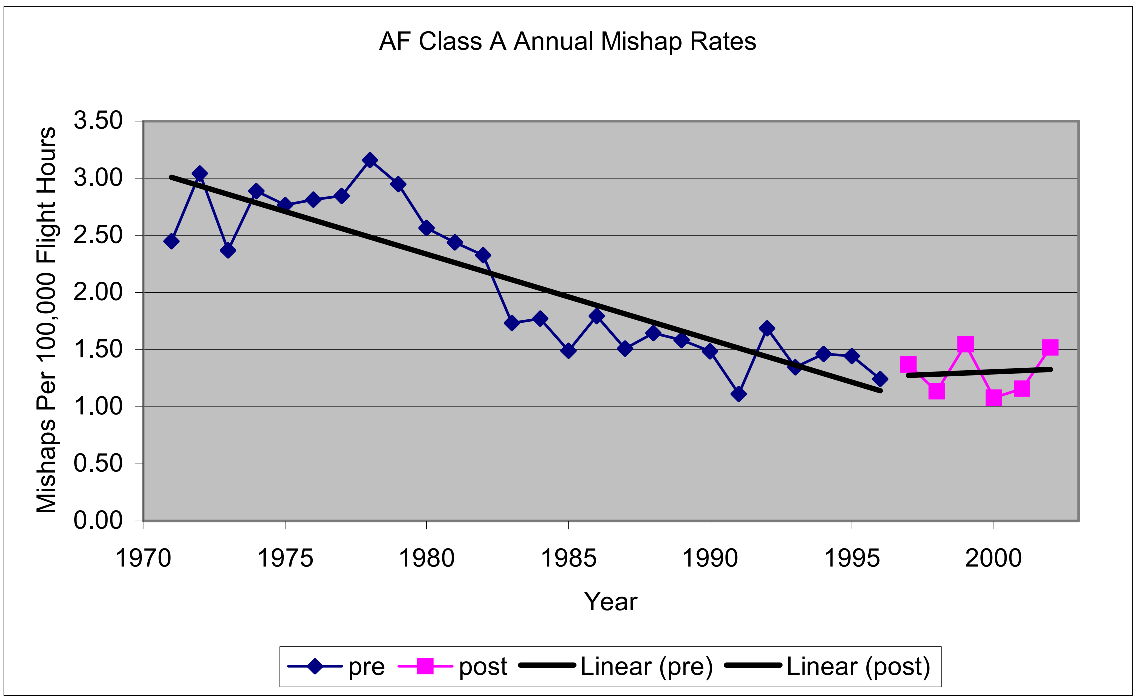


Figure 10. AF Class A Annual Mishap Rates

Tables 15 and 16 show the results of the discontinuous piecewise linear regression tests for AF Class A annual mishap data.

Table 15. AF Class A Annual Overall F Test-Results

Term	Beta	Beta Coefficient	P-Value	Reject Null?	Equal 0?
Y Intercept	β_0	8.801	0.000	NA	NA
FY	β_1	-0.080	0.000	Yes	No
(FY-96)RM	β_2	0.090	0.216	NA	NA
RM	β_3	0.175	0.562	No	Yes

Table 16. AF Class A Annual Partial F-Test Results

SSE	SSR	F Stat	F Crit	Reject Null?	$\beta_2 = \beta_3 = 0$
2.258	0.69	3.97254	4.25	No	Yes

The overall F-tests indicate that the slope of the pre-ORM line, β_1, is significantly different from zero, and that there was no significant shift at the breakpoint at 1996. The partial F-Test does not reject the null hypothesis that both β_2 and β_3 equal zero. This indicates that β_2 is not significantly different from zero and therefore the line after the breakpoint is not significantly different from the line prior to the breakpoint.

Since there was no shift in the regression line in 1996 and the slopes of the two lines are not significantly different, there is no evidence that the implementation of ORM affected the AF Class A Annual mishap rates.

AF Class A Quarterly Mishap Rates.

The AF Class A quarterly mishap rates are illustrated in Figure 11.

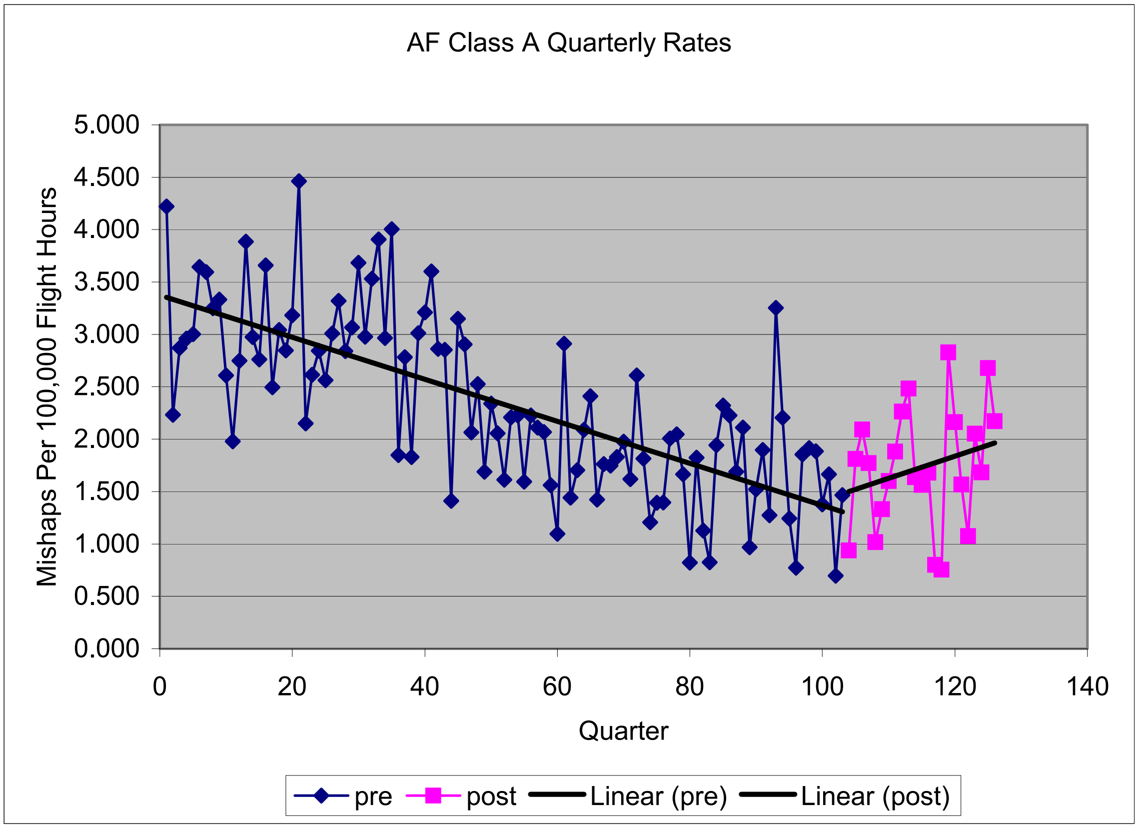


Figure 11. AF Class A Quarterly Mishap Rates

Tables 17 and 18 show the results of the discontinuous piecewise linear regression tests for AF Class A quarterly mishap data.

Table 17. AF Class A Quarterly Overall F Test-Results

Term	Beta	Beta Coefficient	P-Value	Reject Null?	Equal 0?
Y Intercept	β_0	3.376	0.000	NA	NA
FY	β_1	-0.020	0.000	Yes	No
(FY-96)RM	β_2	0.041	0.019	Yes	No
RM	β_3	0.176	0.497	No	Yes

Table 18. AF Class A Quarterly Partial F-Test Results

SSE	SSR	F Stat	F Crit	Reject Null?	$\beta_2 = \beta_3 = 0$
41.663	5.938	8.694	3.800	Yes	No

The overall F-tests indicate that the slope of the pre-ORM line, β_1, is significantly different from zero, and that there was no significant shift at the breakpoint at 1996. The partial F-Test rejects the null hypothesis that both β_2 and β_3 equal zero. Since β_3 was previously shown to equal zero, this indicates that β_2 is significantly different from zero and therefore the line after the breakpoint is significantly different from the line prior to the breakpoint.

While there was no shift in the regression line in 1996, the slopes of the two lines are significantly different, and there is evidence that the implementation of ORM affected the AF Class A Quarterly mishap rates by creating a process change. However, since the slope of the first line is decreasing and the slope of the second line is increasing, it appears ORM did not have its desired effect of reducing rates.

AF Class A Quarterly Sortie Mishap Rates.

The AF Class A quarterly sortie mishap rates are illustrated in Figure 12.

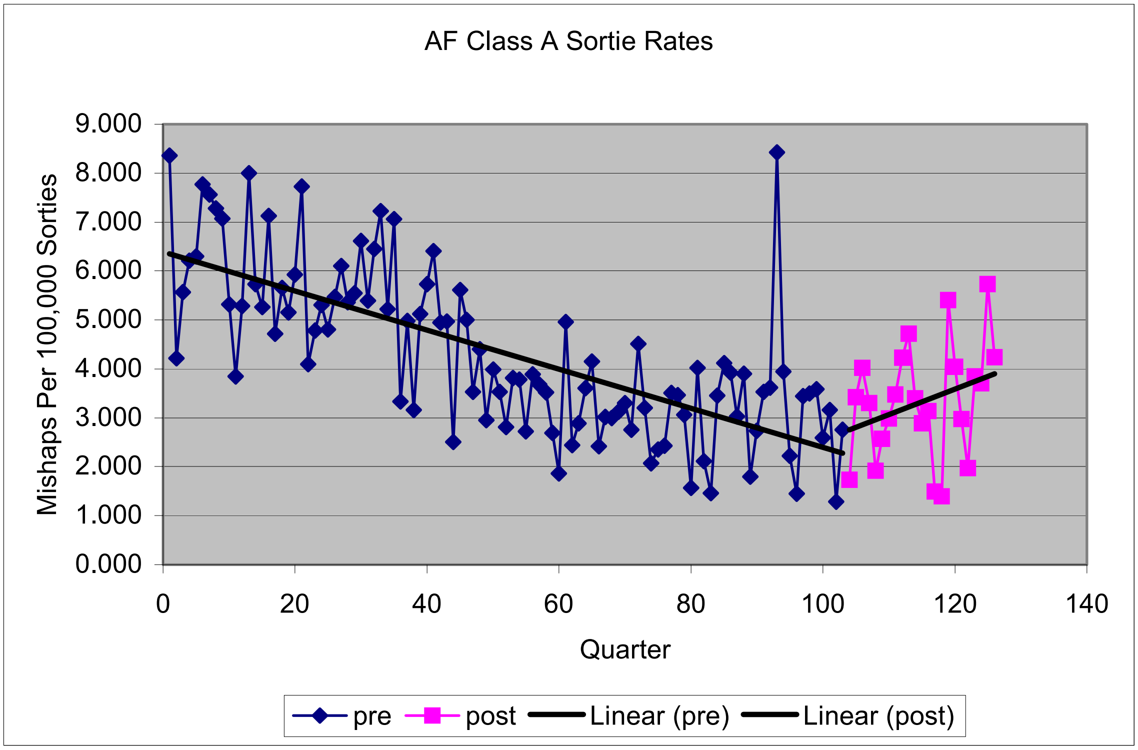


Figure 12. AF Class A Quarterly Sortie Mishap Rates

Tables 19 and 20 show the results of the discontinuous piecewise linear regression tests for AF Class A quarterly sortie mishap data.

Table 19. AF Class A Quarterly Sortie Overall F Test-Results

Term	Beta	Beta Coefficient	P-Value	Reject Null?	Equal 0?
Y Intercept	β_0	6.403	0.000	NA	NA
FY	β_1	-0.040	0.000	Yes	No
(FY-96)RM	β_2	0.092	0.011	Yes	No
RM	β_3	0.460	0.389	No	Yes

Table 20. AF Class A Quarterly Sortie Partial F-Test Results

SSE	SSR	F Stat	F Crit	Reject Null?	$\beta_2 = \beta_3 = 0$
175.425	31.291	10.881	3.800	Yes	No

The overall F-tests indicate that the slope of the pre-ORM line, β_1, is significantly different from zero, and that there was no significant shift at the breakpoint at 1996. The

partial F-Test rejects the null hypothesis that both β_2 and β_3 equal zero. Since β_3 was previously shown to be equal to zero, β_2 is therefore shown to be not equal to zero. This indicates that the line after the breakpoint is significantly different from the line prior to the breakpoint.

While there was no shift in the regression line in 1996, the slopes of the two lines are significantly different. We may therefore conclude that the implementation of ORM affected quarterly sortie mishap rates by creating a process change. However, because the slope of the first line is decreasing (negative) and the slope of the second line is increasing (positive), we cannot conclude that ORM had the desired affect of reducing rates.

AF Class A Operational Causes.

This final Class A analysis examines only operational causes, or mishaps caused by pilot related factors only. Figure 13 shows the number of Class A mishaps with operational causes since 1991.

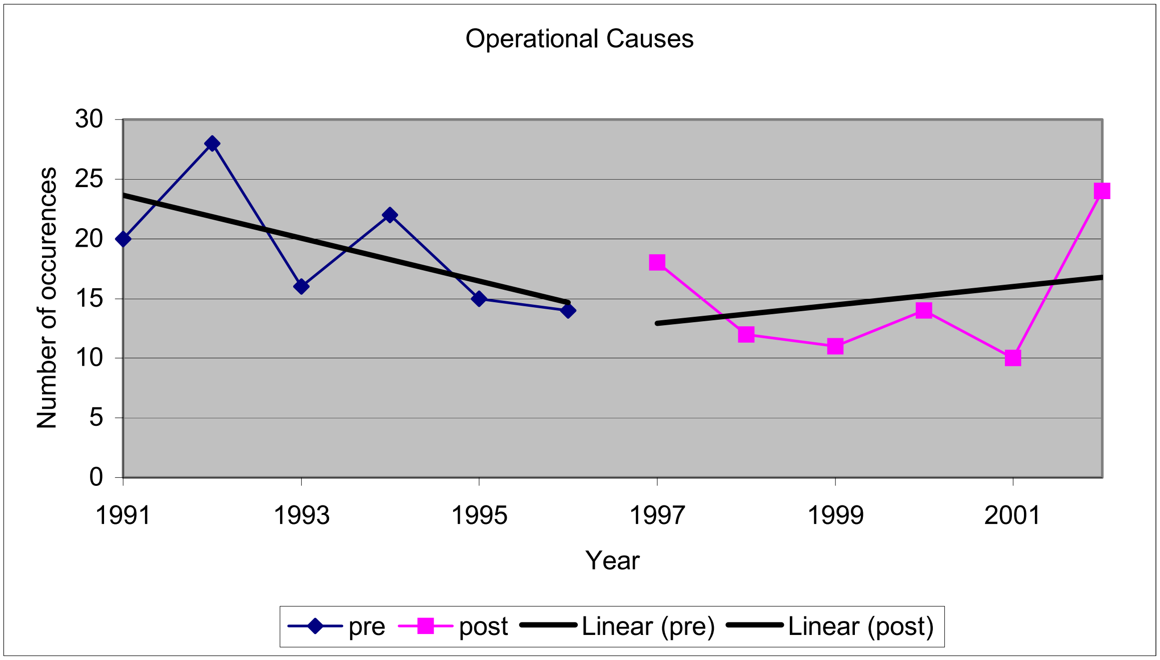


Figure 13. AF Class A Operational Causes

Tables 21 and 22 show the results of the discontinuous piecewise linear regression tests.

Table 21. AF Class A Operational Causes Overall F-Test Results

Term	Beta	Beta Coefficient	P-Value	Reject Null?	Equal 0?
Y Intercept	β_0	3607.467	0.182	NA	NA
FY	β_1	-1.800	0.184	No	Yes
(FY-96)RM	β_2	2.571	0.180	No	Yes
RM	β_3	-2.533	0.689	No	Yes

Table 22. AF Class A Quarterly Sortie Partial F-Test Results

SSE	SSR	F Stat	F Crit	Reject Null?	$\beta_2 = \beta_3 = 0$
241.552	59.000	1.10	6.06	No	Yes

The overall F-tests indicate that the slope of the pre-ORM line, β_1, is not significantly different from zero, and that there was no significant shift at the breakpoint at 1996. The partial F-Test does not reject the null hypothesis that both β_2 and β_3 equal

zero. This indicates that the line after the breakpoint is not significantly different from the line prior to the breakpoint.

There was no shift in the regression line in 1996, and although the pre-ORM line decreased and the pot-ORM line increased, the slopes of the two lines are not significantly different. We may therefore conclude that the implementation of ORM did not affect the number of operational mishap causes. Moreover, because the slope of the first line is decreasing (negative) and the slope of the second line is increasing (positive), we cannot conclude that ORM had the desired affect of reducing rates.

AF Class B Annual Mishap Rates.

The AF Class B annual mishap rates are illustrated in Figure 14. It is interesting to note the substantial spike in Class B mishaps between 1976 and 1979. It is unclear what caused the dramatic increase, which is mirrored in both the quarterly mishap and sortie analysis. Analysis of the period showed no substantial changes in flying hours or sorties flown. There were no major air campaigns being conducted at the time, with the Vietnam War ending in 1975. The most substantial aircraft type suffering Class B mishaps at the time were F-4 variants, but it was not noticeably different than adjacent years. More detailed research using more in-depth data would need to be conducted to learn more about the spike.

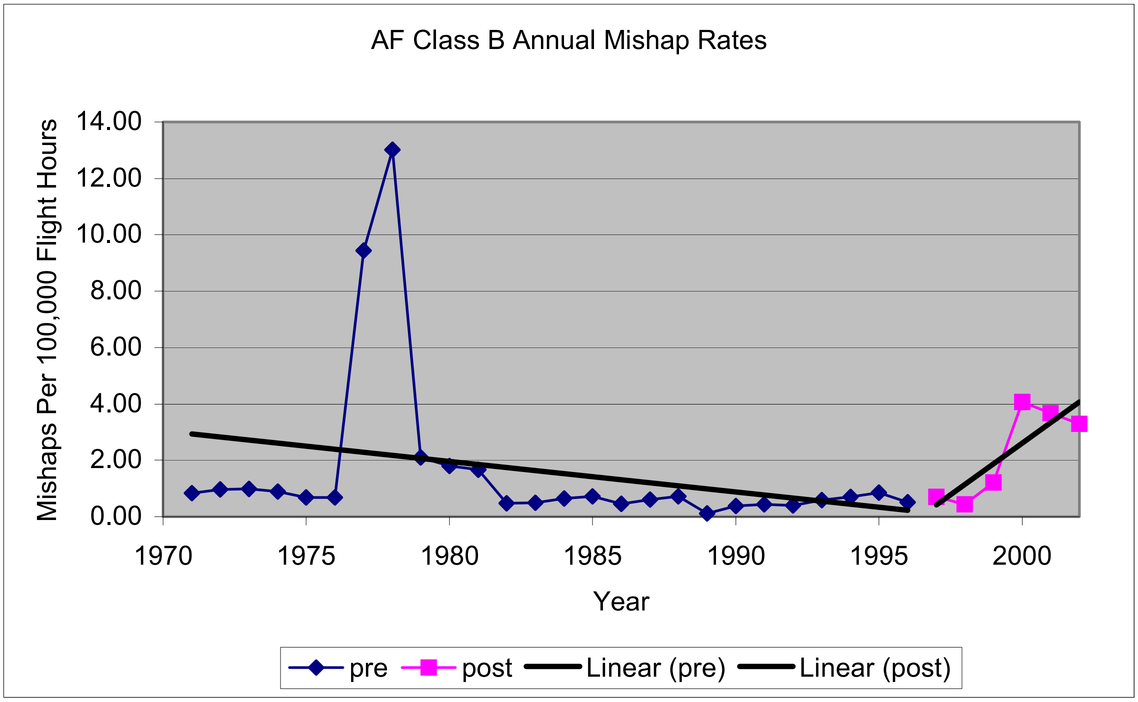


Figure 14. AF Class B Annual Mishap Rates

Table 23 and 24 show the results of the discontinuous piecewise linear regression tests for AF Class B annual mishap data.

Table 23. AF Class B Annual Overall F Test-Results

Term	Beta	Beta Coefficient	P-Value	Reject Null?	Equal 0?
Y Intercept	β_0	14.565	0.040	NA	NA
FY	β_1	-0.153	0.066	No	Yes
(FY-96)RM	β_2	0.883	0.186	No	Yes
RM	β_3	-0.201	0.942	No	Yes

Table 24. AF Class B Annual Partial F-Test Results

SSE	SSR	F Stat	F Crit	Reject Null?	$\beta_2 = \beta_3 = 0$
189.164	31.510	2.165	4.250	No	Yes

The overall F-tests indicate that the slope of the pre-ORM line, β_1, is not significantly different from zero, and that there was no significant shift at the breakpoint

at 1996. The partial F-Test does not reject the null hypothesis that both β_2 and β_3 equal zero. This indicates that β_2 is not significantly different from zero and therefore the line after the breakpoint is not significantly different than the line prior to the breakpoint.

Since there was no shift in the regression line in 1996 and the slopes of the two lines are not significantly different, there is no evidence that the implementation of ORM affected the AF Class B Annual mishap rates.

AF Class B Quarterly Mishap Rates.

The AF Class B quarterly mishap rates are illustrated in Figure 14.

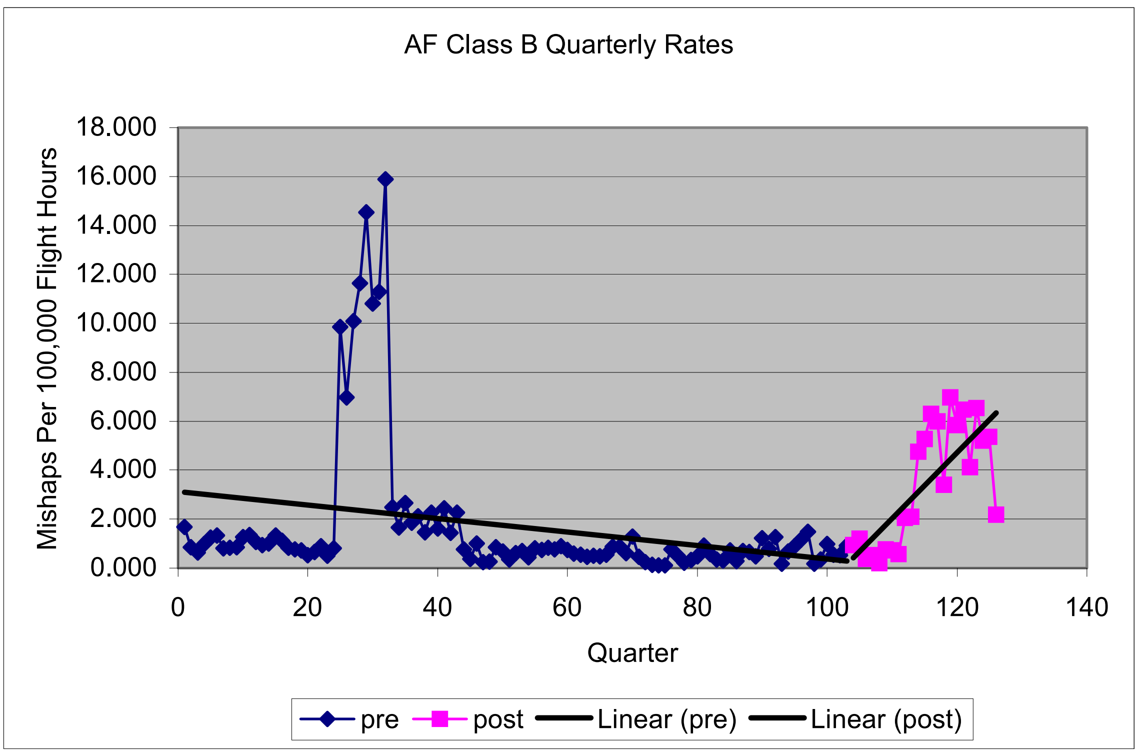


Figure 15. AF Class B Quarterly Mishap Rates

Tables 25 and 26 show the results of the discontinuous piecewise linear regression tests for AF Class B quarterly mishap data.

Table 25. AF Class B Quarterly Overall F Test-Results

Term	Beta	Beta Coefficient	P-Value	Reject Null?	Equal 0?
Y Intercept	β_0	3.133	0.000	NA	NA
FY	β_1	-0.028	0.003	Yes	No
(FY-96)RM	β_2	0.290	0.000	Yes	No
RM	β_3	0.018	0.988	No	Yes

Table 26. AF Class B Quarterly Partial F-Test Results

SSE	SSR	F Stat	F Crit	Reject Null?	$\beta_2 = \beta_3 = 0$
887.254	196.962	13.541	3.800	Yes	No

The overall F-tests indicate that the slope of the pre-ORM line, β_1, is not significantly different from zero, and that there was no significant shift at the breakpoint at 1996. The partial F-Test rejects the null hypothesis that both β_2 and β_3 equal zero. Since β_3 was previously shown to be equal to zero, β_2 is therefore not equal to zero. This indicates that the line after the breakpoint is significantly different from the line prior to the breakpoint.

While there was no shift in the regression line in 1996, the slopes of the two lines are significantly different. We may therefore conclude that the implementation of ORM occurred contemporaneously with an increase in the slope of Class B quarterly mishap rates, indicating a possible process change. Because the slope of the first line is decreasing (negative) and the slope of the second line is increasing (positive), we cannot conclude that ORM had the desired affect of reducing rates.

AF Class B Quarterly Sortie Mishap Rates.

The AF Class B quarterly sortie mishap rates are illustrated in Figure 15.

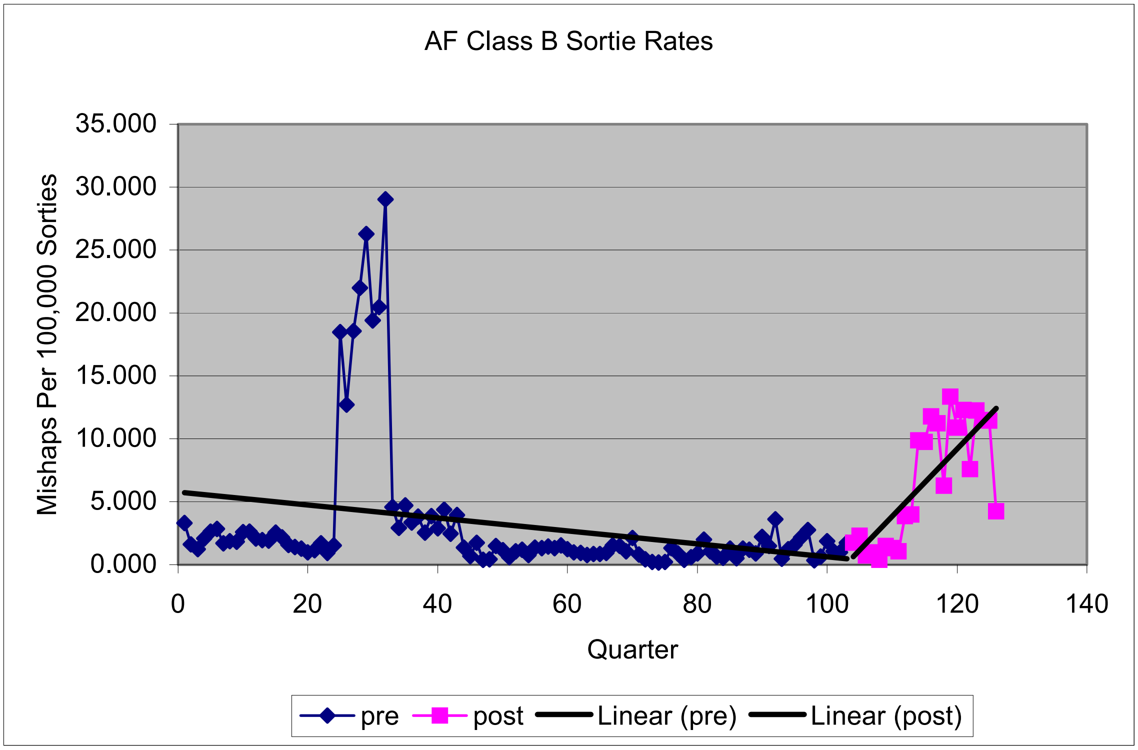


Figure 16. AF Class B Quarterly Sortie Mishap Rates

Tables 27 and 28 show the results of the discontinuous piecewise linear regression tests for AF Class B quarterly sortie mishap data.

Table 27. AF Class B Quarterly Sortie Overall F Test-Results

Term	Beta	Beta Coefficient	P-Value	Reject Null?	Equal 0?
Y Intercept	β_0	5.798	0.000	NA	NA
FY	β_1	-0.052	0.002	Yes	No
(FY-96)RM	β_2	0.572	0.281	No	Yes
RM	β_3	-0.073	0.973	No	Yes

Table 28. AF Class B Quarterly Sortie Partial F-Test Results

SSE	SSR	F Stat	F Crit	Reject Null?	$\beta_2 = \beta_3 = 0$
2976.084	751.462	15.403	3.800	Yes	No

The overall F-tests indicate that the slope of the pre-ORM line, β_1, is significantly different from zero, and that there was no significant shift at the breakpoint at 1996. The partial F-Test rejects the null hypothesis that both β_2 and β_3 equal zero. Since β_3 was previously shown to be equal to zero, β_2 is therefore not equal to zero. This indicates that the line after the breakpoint is significantly different from the line prior to the breakpoint.

While there was no shift in the regression line in 1996, the slopes of the two lines are significantly different. We may therefore conclude that the implementation of ORM occurred contemporaneously with an increase in the slope of Class B quarterly sortie mishap rates, indicating a possible process change. Because the slope of the first line is decreasing (negative) and the slope of the second line is increasing (positive), we cannot conclude that ORM had the desired affect of reducing rates.

AF Class B Quarterly Mishap Rates Revisited.

Earlier analysis of AF Class B annual, quarterly, and sortie rates all indicate increased mishap rates after ORM implementation. A closer examination of the data points revealed an interesting rate surge shortly after the implementation date, starting in 1998. As illustrated in Figure 17 below, the rates remained steady through the ORM implementation quarter (104) and did not begin to increase until July 1998 (quarter 112). The late 1970's rate spike, which appears to be an anomaly of some sort, was avoided during this analysis.

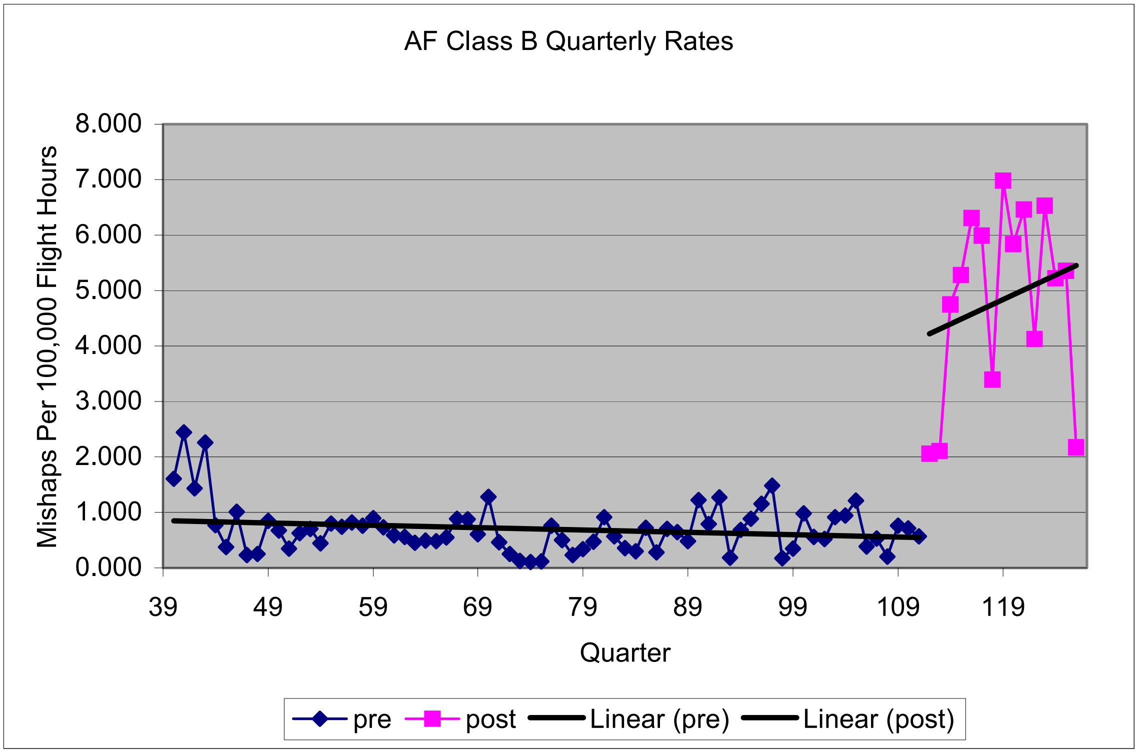


Figure 17. AF Class B Quarterly Mishap Rates Revisited

Examination of the trendlines reveals an obvious shift at the new breakpoint and an obvious increase in the slope of the lines. Tables 29 and 30 show the results of the discontinuous piecewise linear regression tests for AF Class B quarterly sortie mishap data.

Table 29. AF Class B Quarterly ('98) Overall F Test-Results

Term	Beta	Beta Coefficient	P-Value	Reject Null?	Equal 0?
Y Intercept	β_0	1.021	0.004	NA	NA
FY	β_1	-0.004	0.339	No	Yes
(FY-98)Post 98	β_2	0.092	0.055	No	Yes
Post 98	β_3	3.591	0.000	Yes	No

Table 30. AF Class B Quarterly ('98) Partial F-Test Results

SSE	SSR	F Stat	F Crit	Reject Null?	$\beta_2 = \beta_3 = 0$
51.053	132.982	106.796	3.890	Yes	No

Using the resultant p-values, the F-tests indicate that the slope of the pre-1998 line, β_1, is not significantly different from zero and that there was a significant jump at the breakpoint at July 1998. The partial F-Test reveals that β_2 is equal to zero and that the line after 1998 is not significantly different from the line prior to 1998.

There was a significant shift in the regression line in 1998, but the slopes of the two lines are equal. We conclude that some, unexplained event occurred contemporaneously with an increase in the slope of Class B quarterly sortie mishap rates, indicating a process change. The slope of the first line is decreasing (negative) and the slope of the second line is increasing (positive). These results may indicate events other than ORM are causing the observed rate changes.

It is unclear why the AF Class B mishap rates suffered significant increases since 1998. The DoD changed its classification criteria slightly, but that occurred in 2002 and involved only Class C mishaps. It is possible that a surge in operations tempo due to the Operation Allied Force in Kosovo led to the increases, but no such increases are evident during the Gulf War or the Vietnam War. It is also possible that ORM has had a detrimental effect on safety.

Army Results.

Army Class A Annual Results.

Flying hours and sorties flown data was not available so no quarterly data sets were developed. The Army Class A annual mishap rates are illustrated in Figure 18.

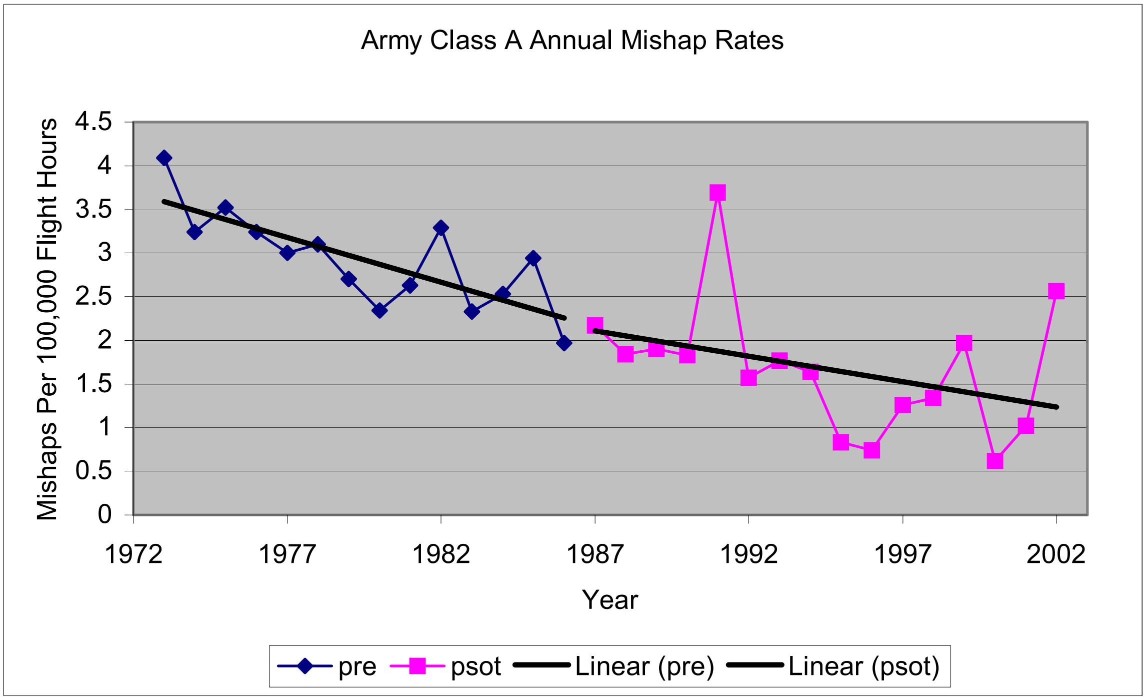


Figure 18. Army Class A Annual Mishap Rates

Tables 31 and 32 show the results of the discontinuous piecewise linear regression tests for Army Class A annual mishap data. Ashley noted very similar numbers in his analysis of the same data ending in 1999 (Ashley, 1999). The only noticeable differences being that Ashley's post-RM slope was slightly steeper and his breakpoint at C was slightly smaller. These differences were fueled by a sharp rate increase after 2000.

Table 31. Army Class A Annual Overall F Test-Results

Term	Beta	Beta Coefficient	P-Value	Reject Null?	Equal 0?
Y Intercept	β_0	11.016	0.001	NA	NA
FY	β_1	-0.102	0.008	Yes	No
(FY-96)RM	β_2	0.045	0.373	No	Yes
RM	β_3	-0.116	0.794	No	Yes

Table 32. Army Class A Annual Partial F-Test Results

SSE	SSR	F Stat	F Crit	Reject Null?	$\beta_2 = \beta_3 = 0$
9.169	0.294	0.417	4.250	No	Yes

The overall F-tests indicate that the slope of the pre-RM line, β_1, is significantly different from zero, and that there was no significant shift at the breakpoint at 1996. The partial F-Test does not reject the null hypothesis that both β_2 and β_3 equal zero. This indicates that β_2 is not significantly different from zero and therefore the line after the breakpoint is not significantly different from the line prior to the breakpoint.

Since there was no shift in the regression line in 1996 and the slopes of the two lines are not significantly different, there is no evidence that the implementation of RM affected the Army Class A Annual mishap rates

Army Class B-C Annual Results.

The Army Class B-C annual mishap rates are illustrated in Figure 19.

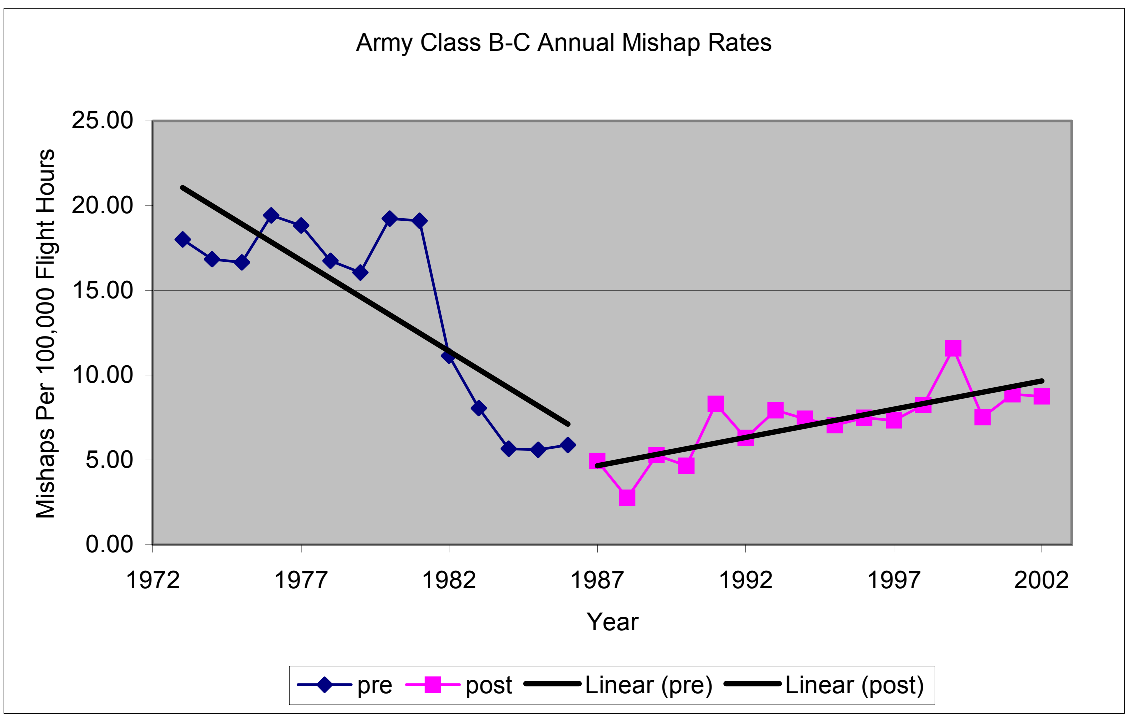


Figure 19. Army Class B-C Annual Mishap Rates

Tables 33 and 34 show the results of the discontinuous piecewise linear regression tests for Army Class B annual mishap data. As with the Class A data, these results are very similar to Ashley's findings (Ashley, 1999).

Table 33. Army Class B Annual Overall F Test-Results

Term	Beta	Beta Coefficient	P-Value	Reject Null?	Equal 0?
Y Intercept	β_0	112.543	0.000	NA	NA
FY	β_1	-1.202	0.000	Yes	No
(FY-96)RM	β_2	1.488	0.000	Yes	No
RM	β_3	-2.768	0.160	No	Yes

Table 34. Army Class B Annual Partial F-Test Results

SSE	SSR	F Stat	F Crit	Reject Null?	$\beta_2 = \beta_3 = 0$
172.672	310.562	23.381	4.250	Yes	No

The overall F-tests indicate that the slope of the pre-RM line, β_1, is significantly different from zero, and that there was no significant shift at the breakpoint at 1987, when the Army officially implemented the program. The partial F-Test rejects the null hypothesis that both β_2 and β_3 equal zero. Since β_3 was previously shown to be equal to zero, β_2 is therefore not equal to zero. This indicates that the line after the breakpoint is significantly different from the line prior to the breakpoint.

While there was no shift in the regression line in 1987, the slopes of the two lines are significantly different. We may therefore conclude that the implementation of RM affected Army Class B annual mishap rates by creating a process change. However, because the slope of the first line is decreasing (negative) and the slope of the second line is increasing (positive), we cannot conclude that RM had the desired affect of reducing rates.

Regressions of Implementation Period.

The AF initiated the implementation of the ORM program in September 1996. That date was therefore chosen to be the breakpoint for the AF regression analyses. It was noted earlier, however, that complete implementation of the program throughout the AF via computer training, was not accomplished until October 1998. It is possible that any reduction in rates due to ORM would not be realized until training was complete.

For this reason, another set of tests was performed on the AF data, this time with two breakpoints; one at September 1996 and another at October 1998. This effectively broke the data set up into three sections; pre-ORM, training, and post-ORM. The same discontinuous piecewise linear regression techniques were used on AF Class A and B quarterly mishap rates.

AF Class A Implementation Period Results.

AF Class A quarterly mishap rate data, segmented into the three periods are shown in Figure 20.

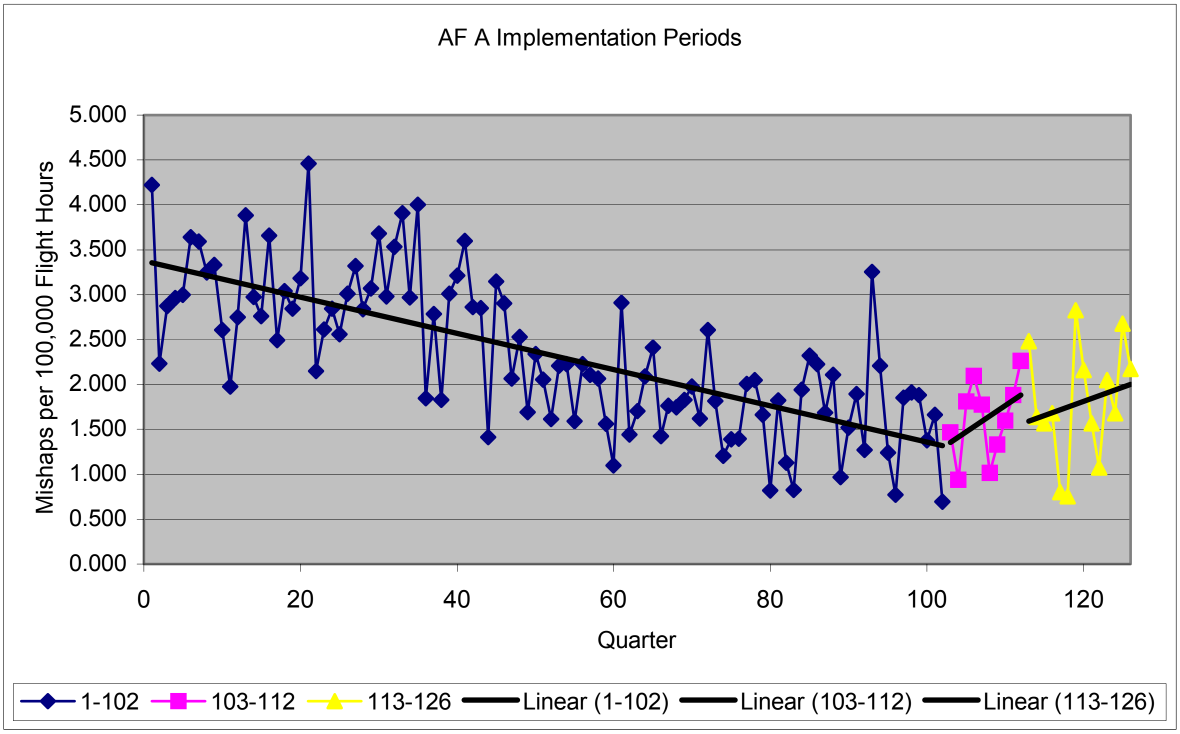


Figure 20. AF Class A Implementation Period Quarterly Rates

Table 35. AF Class A Implementation Period Quarterly Results

	β_3	P-Value	Reject	Equal 0?	F-Stat	F-Crit	Reject
Period 1-2	-0.0207	0.948	No	Yes	2.741	4.25	No
Period 2-3	-0.032	0.503	No	Yes	0.262	3.8	No

Results from the tests, shown in Table 35, reveal that there were no significant shifts at either breakpoint and that the three lines did not have significantly different slopes. It is clear upon examining the chart that the pre-ORM period had a downward sloping rate and that the implementation period had an upward sloping rate. Partial F-Tests on the data set revealed that while the implementation line did begin to increase, the difference was not statistically significant. The lack of significance of the slope shifts are attributed to the relatively weaker strength of the smaller number of data points in the latter periods. The rate line decreased from the implementation to post-ORM periods,

possibly indicating that the benefits of the program were starting to take effect. However, overall, both rates were still increasing.

AF Class B Implementation Period.

AF Class B quarterly mishap rate data, segmented into the three periods are shown in Figure 21.

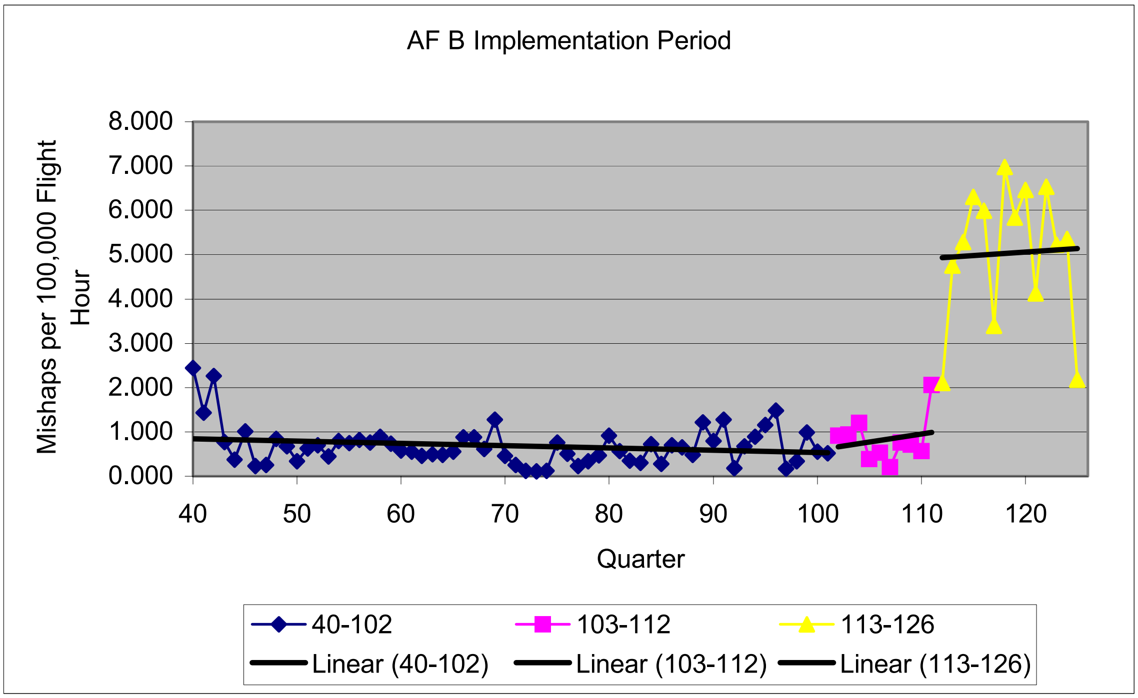


Figure 21. AF Class B Implementation Period Quarterly Rates

Table 36. AF Class B Implementation Period Quarterly Results

	β_3	P-Value	Reject	Equal 0?	F-Stat	F-Crit	Reject
Period 1-2	0.0946	0.775	No	Yes	1.657	3.72	No
Period 2-3	0.103	0.001	Yes	No	7.167	4.32	Yes

Test results, shown in Table 36, indicated that while there was no significant jump at the first breakpoint, there was one at the second, which is evident by the dramatic shift on the chart at quarter 113. The slopes of the three periods, however, were not

significantly different from each other. The implementation period slope increases slightly from the pre-ORM period, and the post-implementation period slope decreases from the implementation period. Although the slope decreased from the previous period, it was still increasing, but it may indicate some successful effects from ORM.

Summary.

To determine whether any rate changes were caused by the implementation of ORM, discontinuous piecewise linear regression was performed. Discontinuous piecewise linear regression determines whether a slope or intercept change is present at a selected point in time (Neter and others, 1996). Analyses were performed on a number of data sets, including annual rates, quarterly rates, quarterly sortie rates, and human factors mishaps. Class A and B data were analyzed for both the Air Force and the Army. None of the tests indicated a downward shift in mishaps nor a reduction in slope after implementation.

IQ.5: Have the proportion of human factor related mishaps changed since implementation?

To determine whether the relative proportion of human factor mishap causes have changed since implementation of ORM, the Chi-Square Goodness of Fit Test was used to analyze annual causal data. If ORM were effective, one would expect a reduction in the proportion of human factor mishaps. A summary of the Chi-Square test results can be found in Appendix S. The overall proportions of AF human factor causes since 1991 are shown in Figure 22. Human factors play a role in about 70% of Class A mishaps and 40% of Class B mishaps.

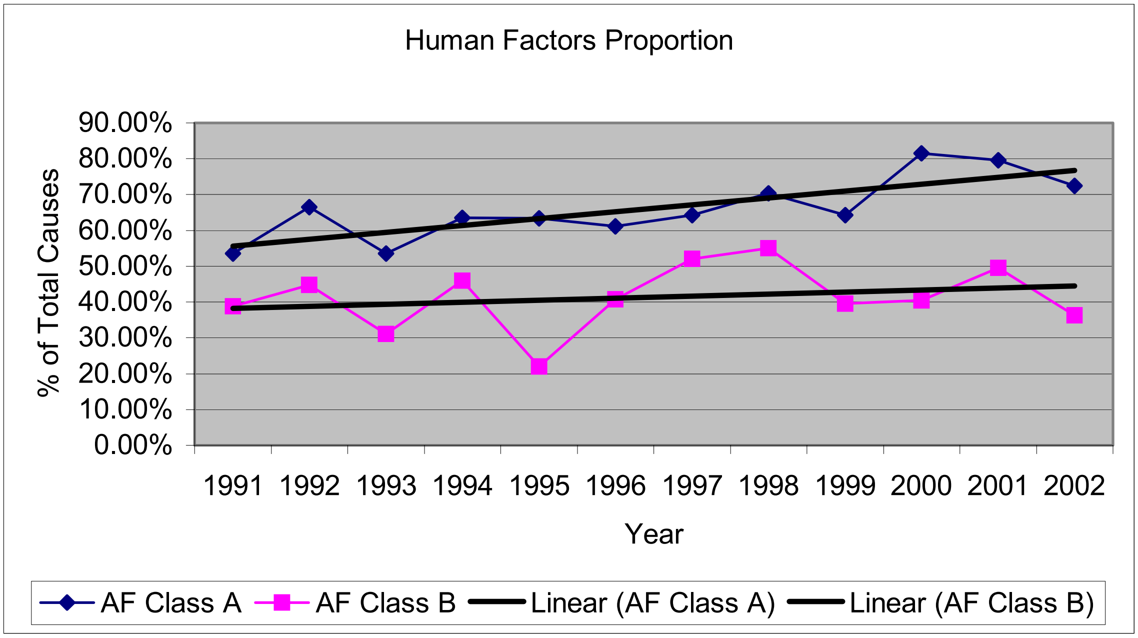


Figure 22. AF Human Factors Mishap Proportions

The trend line for Class A mishaps increased slightly, while the trend line for Class B mishaps remained nearly level.

The overall proportions of Army human factors causes are shown in Figure 23. Approximately 80% of Army aviation mishaps are human factors related.

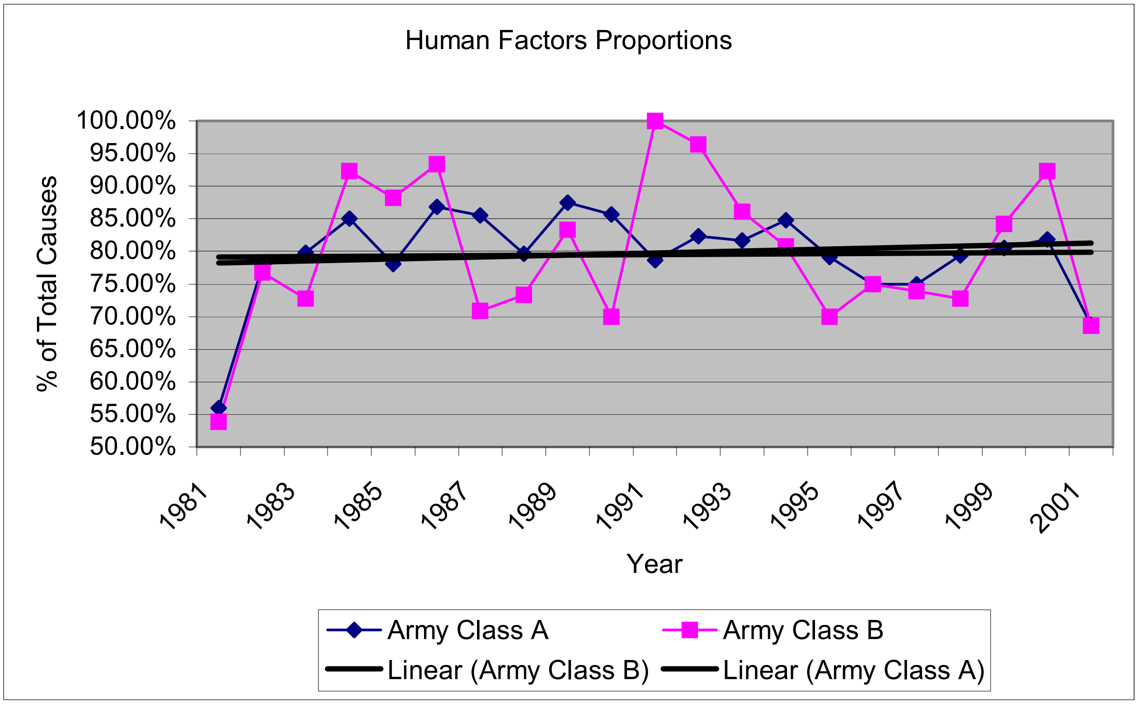


Figure 23. Army Human Factors Mishap Proportions

Data.

AF and Army Class A and B annual mishap causal data were analyzed in these tests. The data format was a yearly number count of mishap categories provided by analysts from both the AF and Army Safety Centers (Air Force Safety Center, 2003b), (Army Safety Center, 2002). Mishaps may have more than one associated cause category.

AF Results.

The Chi-Squared test identified differences in the sample behavior of pre- and post-ORM mishap cause categories. The causes examined, their hypothesized proportion, observed frequencies, and expected frequencies for Class A are shown in Table 37.

Table 37. AF Class A.1 Chi-Square Values

	Hyp Prop	Freq	Exp Freq	Inc/Dec
Accepted Risk	0.03	38	19.90	Inc
Attention Mgt	0.00	49	1.00	Inc
Cognitive Funct	0.00	21	1.00	Inc
Discipline	0.04	18	27.48	Dec
Emotional State	0.09	40	54.97	Dec
Inad Risk Assess	0.01	21	4.74	Inc
Judgment	0.25	135	160.16	Dec
Manning	0.00	1	1.00	--
Perceptions	0.07	47	44.54	--
Physiological	0.03	9	18.01	Dec
Preparations	0.01	14	9.48	Inc
Proficiency	0.03	25	17.06	Inc
Self Induced Stressors	0.00	2	1.00	--
Training	0.04	17	23.69	Dec
Unauth Mod	0.00	1	1.00	--
Unknown	0.12	43	75.81	Dec

Given a 0.05 level of significance and 15 degrees of freedom, the critical χ^2 value was 32.8. The test statistic, χ^2, using the data from Table 37 equaled 2787.57, well beyond the critical value and well inside the rejection region.

Based on these results, the null hypothesis that the proportions are the same is rejected, indicating that a statistically significant change had occurred since implementation. Interestingly, the Accepted Risk and Inadequate Risk Assessment categories both showed significant increases, which is counter-intuitive to expectations. Several of the categories had extremely large increases because there were no recorded incidences of those categories before 1996. A shift in the accounting policy is likely, so the test was run again with those categories removed. These values are shown in Table 38.

Table 38. AF Class A.2 Chi Squared Values

	Hyp Prop	Freq	Exp Freq	Inc/Dec
Accepted Risk	0.03	38	12.78	Inc
Discipline	0.04	18	17.64	Dec
Emotional State	0.09	40	35.29	Dec
Inad Risk Assess	0.01	21	3.04	Inc
Judgment	0.25	135	106.5	Dec
Perceptions	0.07	47	28.59	--
Physiological	0.03	9	11.56	Dec
Preparations	0.01	14	6.08	Inc
Proficiency	0.03	25	10.95	Inc
Training	0.04	17	15.21	Dec
Unknown	0.12	43	48.67	Dec

With 10 degrees of freedom, the critical χ^2 value is 18.3. The resultant χ^2 value is 110.2, far exceeding the critical value and indicating a rejection of the null hypothesis. We may therefore conclude that the population proportions are different. While five of the categories decreased, another five exceeded their expected frequencies, contrary to expectations. One category, judgment, maintained its historical proportions. It cannot, therefore, be concluded that the proportions of human factor related Class A mishaps have decreased since ORM was implemented. Accepted Risk and Inadequate Risk Assessment once again increased in proportions.

Results from the Chi-Squared test on AF Class B data are shown in Table 39.

Table 39. AF Class B Chi Squared Values

	Hyp Prop	Freq	Exp Freq	Inc/Dec
Accepted Risk	0.07	21	30.28	Dec
Attention Mgt	0.00	17	1.00	Inc
Cognitive Funct	0.00	5	1.00	Inc
Discipline	0.05	9	20.19	Dec
Emotional State	0.04	14	17.66	Dec
Inad Risk Assess	0.01	17	2.52	Inc
Judgment	0.08	58	35.33	Dec
Manning	0.01	1	1.00	--
Perceptions	0.06	17	27.76	Dec
Physiological	0.01	0	2.52	Dec
Preparations	0.02	6	10.09	Dec
Proficiency	0.00	11	0.00	Inc
Self Induced Stressors	0.00	0	1.00	--
Training	0.02	12	7.57	Inc
Unauth Mod	0.00	0	1.00	--
Unknown	0.15	53	65.60	Dec

Given a 0.05 level of significance and 15 degrees of freedom, the critical χ^2 value was 32.8. The test statistic, χ^2, using the data from Table 39, equaled 379.41, well beyond the critical value and well inside the rejection region.

Based on these results, the null hypothesis that the proportions are the same is rejected, indicating that a statistically significant change had occurred since implementation. Five of the categories increased their proportions, eight decreased, and three remained relatively unchanged. In general, results show that the human factors causal proportions slightly decreased since ORM implementation. ORM specific categories were inconclusive, as the Accepted Risk category showed a decrease while Inadequate Risk Assessment categories increased.

Army Results.

The same testing methodology was applied to Army human factors causes. A shift in accounting methods and terminology in 1995 yielded any such data beyond that point unreliable. Therefore, the Army tests include only data from 1981 to 1994, seven years after implementation. Class A mishap categories, hypothesized proportions, observed frequencies, and expected frequencies are shown in Table 40. Class B values are shown in Table 41.

Table 40. Army Class A Chi-Square Values

	Hyp Prop	Freq	Exp Freq	Inc/Dec
Anticipate	0.04	16	16.41	--
Comply w/ General Rules	0.05	19	20.32	--
Follow Procedures/Orders	0.15	73	56.26	Inc
Recognize	0.05	18	20.32	--
Attention	0.08	21	31.26	Dec
Complex Physical Action	0.09	25	35.16	Dec
Decision	0.11	30	40.64	Dec
Communication	0.03	30	10.94	Inc
Inspection/Search	0.04	42	15.63	Inc
Planning	0.06	5	22.66	Dec
Insufficient Info Reported	0.00	3	1.56	--
Misinterpreted	0.01	8	4.69	Inc
Clearance/Speed/Weight	0.04	22	16.41	Inc

Table 41. Army Class B Chi-Square Values

	Hyp Prop	Freq	Exp Freq	Inc/Dec
Anticipate	0.05	17	7.89	Inc
Comply w/ General Rules	0.03	8	3.94	Inc
Follow Procedures/Orders	0.15	52	22.88	Inc
Recognize	0.08	25	11.83	Inc
Attention	0.02	21	2.37	Inc
Complex Physical Action	0.10	28	14.99	Inc
Decision	0.09	32	13.41	Inc
Communication	0.07	33	11.05	Inc
Inspection/Search	0.05	20	7.10	Inc
Planning	0.06	23	9.47	Inc
Insufficient Info Reported	0.01	4	1.58	Inc
Misinterpreted	0.02	11	3.16	Inc
Clearance/Speed/Weight	0.08	24	11.83	Inc

Given a level of significance of 0.05 and 12 degrees of freedom, the χ^2 critical value is 28.29 for both tests. The Class A data yielded a χ^2 of 111.46. The Class B χ^2 was 372.29. Both tests yielded χ^2 values that exceeded the critical value, so both tests reject the null hypothesis. Therefore, we may conclude that the population proportions are not equal. For Class A data, it is inconclusive whether proportions have gone down. For Class B, all proportions increased. Results of the tests can be found in Appendix S.

Summary.

Using the Chi-Squared Goodness of Fit Test, this investigative question sought to determine whether the proportions of human factor related mishaps had decreased since the implementation of ORM. Since ORM is designed to assist individuals in their decision-making and risk assessment skills, one would expect to see reductions in risk specific cause categories and in human factor cause categories in general.

The AF data revealed evidence to the contrary. While both Class A and B showed significant changes in human factor proportions, neither showed significant decreases. Class B proportions slightly decreased while Class A remained even. Alarmingly, Class A risk specific categories actually increased proportions.

The Army likewise showed significant changes in both Class A and Class B data. Class A proportions did not conclusively increase or decrease, but Class B causes increased unanimously.

Summary

The purpose of this chapter was to answer the overall research question by answering the five investigative questions posed in Chapter 1. For each investigative

question the problem was restated, relevant data was described, and answers were presented according to the methodology described in chapter 3.

The analysis of investigative questions 3 and 4 ultimately allowed us to identify differences in the mishap rates contemporaneous with RM and ORM implementation. None of the data sets analyzed, for the Air Force or the Army, conclusively showed a decrease in mishap rates occurring contemporaneously with ORM implementation, and several showed significant increases. Investigative question 5 identified that the changes were also contemporaneous with changes in the proportion human factors mishap causes of all four data sets analyzed, and with increases in three of the four. The results of the questions provide strong circumstantial evidence that ORM and RM did not cause reductions in mishap rates and that it may be associated with any decreases or increases.

The results of the experiment do not identify a causal relationship between the mishap rate increases and the ORM program. Instead, they suggest that the changes seem to occur at a point in time concurrent with the implementation of ORM, which could, in fact, be attributable to a number of other factors, such as sample demographics, aircraft mix, operational changes due to contingency involvement, social turbulence, and others. Further discussion of these possible confounds can be found in Chapter 5.

Chapter V

Chapter Overview

Despite having drastically reduced its mishap rates since the birth of the AF in 1947, the need to eliminate mishaps altogether persists. In an effort to continue reducing mishap rates, the Air Force implemented the ORM program in 1996, emphasizing an atmosphere of safety at all levels.

A study of the Army's RM program, the model for the Air Force's ORM, was conducted in 1999, revealing that the program failed to significantly improve Army aviation mishap rates (Ashley, 1999). In fact, the findings of the research suggested that accident rates actually increased after RM implementation. The study concluded that the Air Force should therefore not expect mishap rates to decline due to implementation of their ORM program.

This research effort follows Ashley's recommendation that the AF ORM program and its effects on aviation mishap rates be studied. Its research objective was to determine to what degree the implementation of ORM has affected flying safety in the Air Force.

This chapter reviews the findings of the research based on the answers of the investigative questions. It then presents the final conclusions of the thesis by answering the overall research question. Recommendations for the AF's and Army's future use of ORM are made based on the findings. Finally, proposals for future research of topics stemming from the thesis are presented.

The AF acknowledges that recent mishap trends have not been positive. After calling for increased participation in the ORM program in June 2002 (Jumper, 2002a), AF Chief of Staff, General John J. Jumper readdressed the issue describing a disturbing trend in aviation mishaps in a 20 December 2002 memorandum. He pointed out that with 33 Class A mishaps and 22 fatalities, it amounted to one Class A every ten days and the equivalent of one entire lost squadron of aircraft valued at $820 million. Jumper stated that 2002 was the third worst in the last ten years in terms of flying safety, and cited human factors as the cause in two-thirds of the accidents. Inexperience, "edge of the envelope" flying, insufficient or inadequate guidance, and procedure deviations were the leading human factor causes (Jumper, 2002b).

General Jumper stated his concern over the increasingly negative trend and pointed out that safety and mission are inseparable. He identified the use of risk management principles to identify hazards and reduce risks as the best means of reversing the trends (Jumper, 2002b).

Findings

ORM was developed and implemented as the AF's primary means of developing a safer AF. It was intended to make flying safer, thereby reducing mishaps. The overall objective of this research was to determine whether or not the program was successful in those endeavors. To that end, a successful Operational Risk Management program should see a number of beneficial changes. First, it should have seen an overall decrease of mishaps after it was implemented. Second, it should have enjoyed an immediate downward shift in mishaps and a decreasing trend in mishap rates. Third, it should have forced a decrease in the relative proportion of mishaps due to human error.

Based on the results of the analysis of the investigative questions, there is enough evidence to say that ORM has not had its desired effect of reducing mishap rates for the AF. Comparison of means testing failed to conclusively show a reduction in mishap rates after implementation of the program. Discontinuous piecewise linear regression failed to show a decreasing shift nor decreasing slopes in any of the mishap data sets analyzed. Chi squared analysis of the human factor mishap proportions showed that a process change had occurred and proportions had changed, but were generally on the rise. Furthermore, risk-specific categories were growing in proportion.

Summary of Confounds

Implementation.

An assumption of this thesis is that ORM has been fully implemented and is being actively used by Air Force pilots. It is possible, however, that it is not being used or that it is being used incorrectly, and under such circumstances, the mishap increases could not be linked causally with ORM. We would nevertheless come to the same conclusions, that ORM was not having its desired affects, although for different reasons.

Social Considerations.

The military was undergoing considerable social turbulence during the mid- to late-nineties, which could have had negative affects on attitudes, moods, and other psychological aspects. Military controversies, such as the court martial of Lt. Kelly Flynn and suicide of Admiral Jeremy Borda, as well as the scandal in the white house were contributors. Since ORM is essentially a social program, it may have been affected by such turbulence, thereby skewing the measurable affects of ORM.

Manpower Demographics.

One noted history threat to the thesis was maintenance and pilot manning. Notable decreases in fill rates for key maintainer positions and pilot retention problems coincide with mishap rate increases in the late 90's and early 2000. A more comprehensive study into manning of these key positions and mishaps needs to be conducted to determine if there is a substantial relationship, and if so, it could play a significant role in the findings, diminishing the lack of affects attributed to ORM.

Aircraft Mix.

A perfunctory examination of the aircraft fleet was performed without indicating any obvious confounds, but a more detailed and in-depth analysis must clearly be conducted. For example, each aircraft type should be studied separately to detect changes and ageing aircraft should be removed or analyzed separately. This thesis analyzed aggregate data, composed of the entire Air Force fleet. Various affects from the aircraft mix diminish the strength of the findings.

Kosovo.

A simple study of mishap rates in several major contingencies revealed that rates tended to go down. However, Air Force Class B mishaps showed a notable increase during involvement in Kosovo. It is possible that austere runway conditions in the region contributed to the increases. Severe Class A mishaps would most likely not be affected, but less severe Class B's, which could include the wear and tear of hard landings and associated damage, might have been increased.

Diminishing Returns.

Chi-squared testing revealed that the Army enjoyed successful reductions in their proportions of risk-related causes after implementation, while the other data sets did not. This particular data set had a high rate of pre-ORM incidents of risk-related mishaps, while the others did not. It might indicate that ORM is effective in reducing risk-related mishaps if they are a serious problem, but not effective if the risk-related mishaps are less historically significant.

Breakpoints.

Significant slope changes were found at the Air Force ORM breakpoint in 1996 and in the Army RM breakpoint in 1987 for Class B-C mishaps. It is possible that some significant event, other than ORM/RM implementation happened at those breakpoints to cause the slope changes. If the Air Force rates showed a similar change in slope at the Army breakpoint, and if the Army rates showed changes at the Air Force breakpoint, it would decrease the likelihood of ORM/RM being responsible for the slope changes. Figures 24 and 25 illustrate the switched breakpoints.

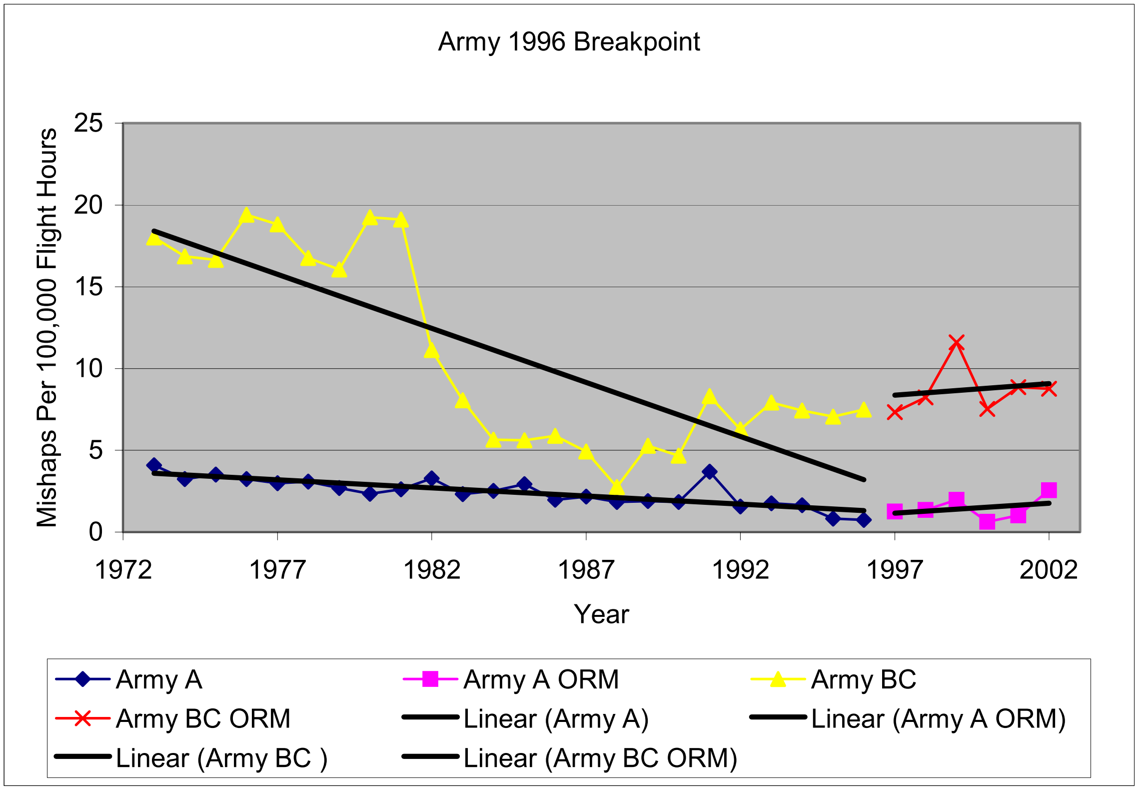


Figure 24. Army Class A and B-C 1996 Breakpoint

The Army implemented its program in 1987 and did not experience any slope reductions at that breakpoint. Using 1996 as a breakpoint, there appears to be a slope increase in Class B-C mishap rates. This may indicate a history threat of some sort in 1996, where the Air Force also detected significant slope increases. If so, it would decrease the likelihood that ORM caused the increases.

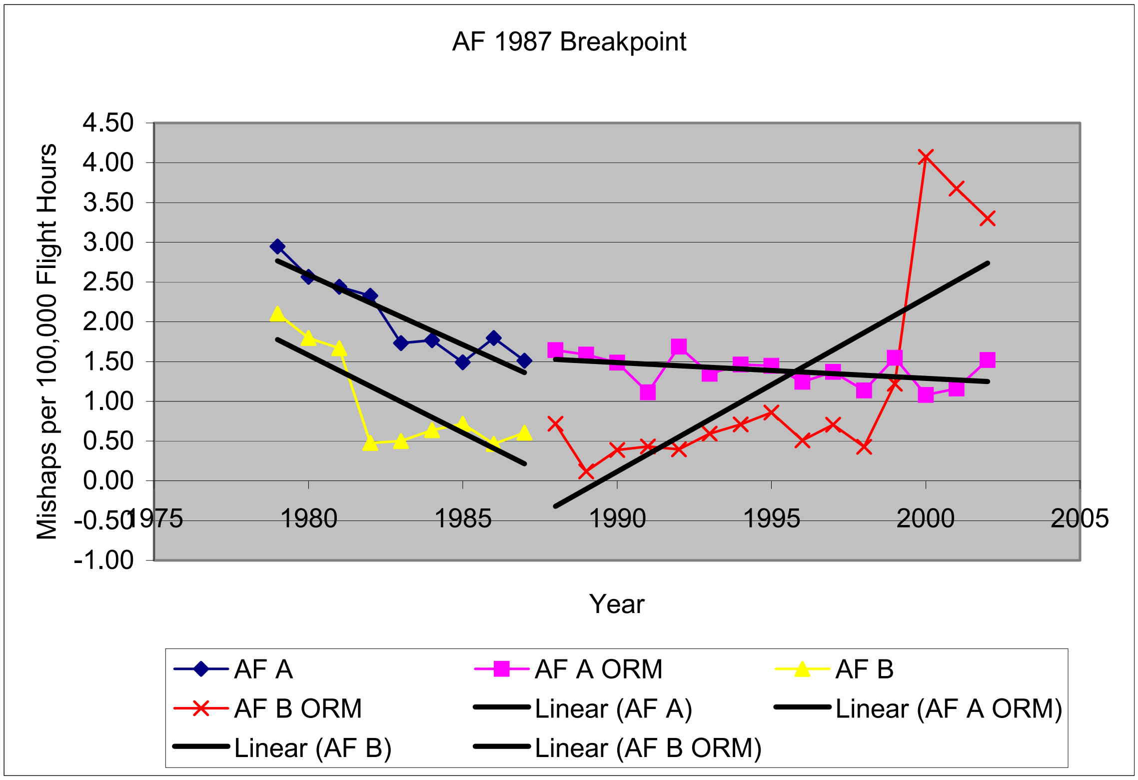


Figure 22. AF Class A and B 1987 Breakpoint

The Army experienced a significant slope increase in Class B mishap rates in 1987, when it implemented its RM program. The Air Force also shows an obvious slope change, which could also indicate that some historical significant event occurred, causing the increases. Since ORM was not implemented until 1996, this would rule it out as a cause. Conversely, the Class B rate appears to hold steady until 1993, and in fact, had a noticeable decrease at 1987.

Recommendations

The findings suggest that ORM was not effectively reducing mishap rates, but were complicated with a number of unexplained and uncontrolled confounds. Therefore the recommendation of this thesis is to conduct more research into the problem. Ideas for further research are described in the next section of this chapter.

Future Research

Several proposals for future research, based on findings and confounds that arose during the research, are now presented.

Proposal 1.

History is often a source of confound within time-series experiments. One such confound may exist in the aircraft mix within the fleet during the analysis. The tests in this thesis were conducted on aggregate data--that is, the entire mix of aircraft within the fleet. Useful insight could be garnered with studies on individual aircraft types. For example, it might be interesting to uncover the mishap rate trends over the lifespan of the F-16 or whether the associated mishap causes were increasingly due to human error or not. The removal from the data pool of the older, retired planes could also affect analytical outcomes. Further analysis of maintainability factors, ageing, and associated safety factors could shed some valuable light on the Air Force's mishap trends.

Proposal 2.

The Navy and Marine Corps began officially implementing similar RM programs in April 1997. A similar study of their mishap data could prove useful to further identify ORM's efficacy. If the Navy/Marine Corps program were to show mishap reductions, understanding what they are doing differently could shed light on the Army and AF's apparent lack of results.

Proposal 3.

This research was conducted on limited data. The AFSC could only provide one type of aviation data for analysis: unsuccessful sorties. To better analyze and understand ORM's effects on sorties, successful sortie data should be included. Failures (mishaps)

and successes (non-crashes) should be analyzed side by side to identify the key factors that cause mishaps. If extensive data were available for each sortie, a multivariate factor analysis could be conducted. Unfortunately, such data is not currently available. A study leading to the development of a comprehensive database that could store all such data would be enormously valuable to the safety community.

Proposal 4.

The AF implemented ORM to establish an atmosphere of safety at all levels at all times, including non-aviation and off-duty activities. Techniques similar to the ones used in this research could be performed on non-aviation mishaps to learn more about the effects of ORM.

Summary

This chapter presented findings, conclusions, recommendations, and future research proposals. The research studied the AF's ORM program and its effects on aviation safety to determine whether ORM had successfully reduced aviation mishap rates. Analysis identified several significant increases in the slopes of mishap rates contemporaneous with the implementation of ORM. It concluded that the AF has not seen a significant reduction in its aviation mishap rates since ORM was implemented and recommends further research.

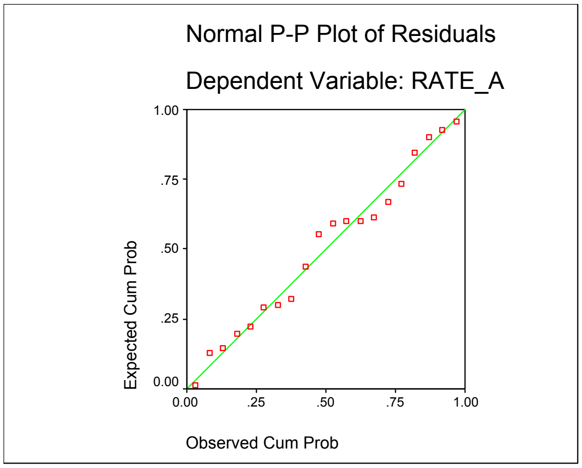
Appendix A. USAF Historical Mishap Data

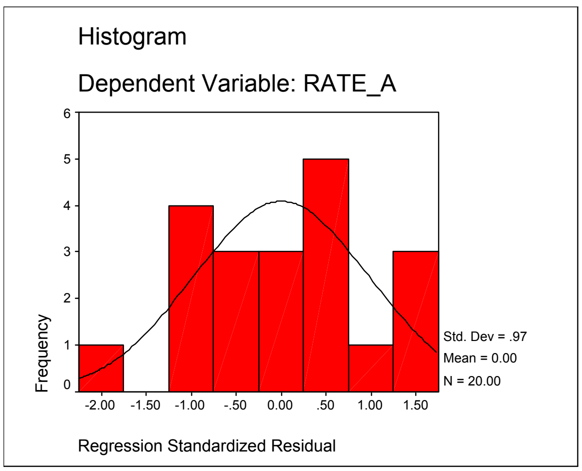
USAF HISTORY					
1947-2002					
	CLASS A		CLASS B		
YEAR	#	RATE	#	RATE	HOURS
65	304	4.57	85	1.28	6648121
66	345	4.91	95	1.35	7030015
67	332	4.54	104	1.42	7311121
68	311	3.90	101	1.27	7983688
69	299	4.05	66	0.89	7388976
70	201	3.05	58	0.88	6597248
71	141	2.45	48	0.83	5754376
72	163	3.04	52	0.97	5356984
73	102	2.37	42	0.98	4307058
74	108	2.89	33	0.88	3736870
75	93	2.77	23	0.68	3359170
76	87	2.81	21	0.68	3094317
77	90	2.84	299	9.45	3164334
78	98	3.16	404	13.02	3102541
79	94	2.95	67	2.10	3189969
80	81	2.56	56	1.80	3118938
81	80	2.44	54	1.67	3237027
82	78	2.33	16	0.48	3352123
83	59	1.73	17	0.50	3407085
84	62	1.77	22	0.64	3446291
85	53	1.49	25	0.72	3491318
86	62	1.79	16	0.46	3454716
87	40	1.51	16	0.60	2650469
88	55	1.64	24	0.72	3345384
89	56	1.59	4	0.12	3406291
90	51	1.49	13	0.39	3366126
91	41	1.11	16	0.43	3684741
92	48	1.69	11	0.39	2787867
93	34	1.35	15	0.59	2525935
94	35	1.46	16	0.71	2256251
95	32	1.44	19	0.86	2215310
96	27	1.24	11	0.51	2169696
97	29	1.37	15	0.71	2119682
98	24	1.14	9	0.43	2111154
99	33	1.55	26	1.22	2130958
00	22	1.08	83	4.08	2036757
01	24	1.16	76	3.68	2067104
02	35	1.52	76	3.30	2303480
TOTAL	3829	2.28	2134	1.62	142709491

Appendix B. US Army Historical Mishap Data

US ARMY HISTORY					
1973-2002					
	Class A		Class B-C		Flying
FY	Number	Rate	Number	Rate	Hours
73	64	4.09	282	18.02	1564594
74	51	3.24	265	16.85	1572314
75	52	3.52	246	16.65	1477625
76	48	3.24	288	19.41	1483553
77	45	3	282	18.81	1498906
78	45	3.1	243	16.76	1449788
79	39	2.7	232	16.07	1443836
80	36	2.34	296	19.25	1537508
81	43	2.63	312	19.11	1632790
82	52	3.29	176	11.14	1580162
83	37	2.33	128	8.05	1589599
84	39	2.53	87	5.65	1538610
85	45	2.94	86	5.61	1531829
86	32	1.97	96	5.90	1628163
87	37	2.17	84	4.93	1704675
88	32	1.84	48	2.76	1741997
89	32	1.9	89	5.28	1685100
90	31	1.83	79	4.67	1690601
91	48	3.69	108	8.31	1299734
92	22	1.57	88	6.29	1400052
93	23	1.77	103	7.93	1299337
94	21	1.64	95	7.43	1278098
95	10	0.83	85	7.06	1203719
96	8	0.74	81	7.49	1082006
97	12	1.26	70	7.35	952999
98	12	1.34	74	8.24	897870
99	18	1.97	106	11.60	913705
00	6	0.619	73	7.54	967741.9
01	10	1.022	87	8.87	980392.2
02	26	2.56	89	8.76	1015625
TOTALS	976	2.2557	4378	10.39	41642929

Appendix C. AF Class A Residual Frequency Distribution and Normality Test





Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1.2266	1.6854	1.4560	.1429	20
Residual	-.3822	.2934	-1.0658E-15	.1680	20

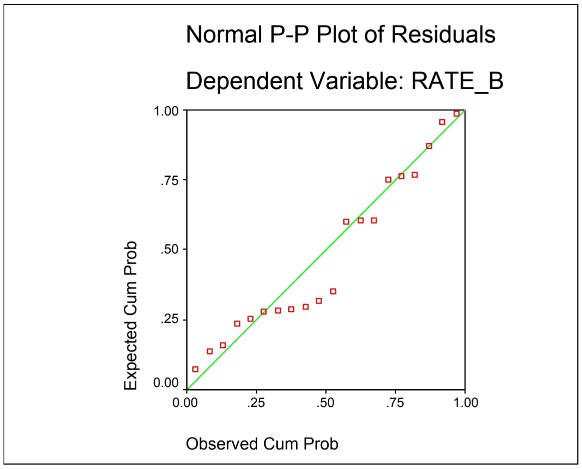
a Dependent Variable: RATE_A

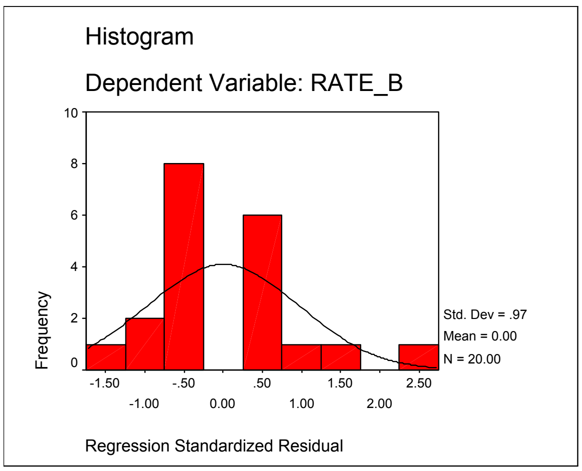
One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
Kolmogorov-Smirnov Z		.464
Asymp. Sig. (2-tailed)		.983

a Test distribution is Normal.
b Calculated from data.

Appendix D. AF Class B Residual Frequency Distribution and Normality Test





Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-.1564	2.2624	1.0530	.7532	20
Residual	-1.3232	2.0722	-9.1926E-15	.8863	20

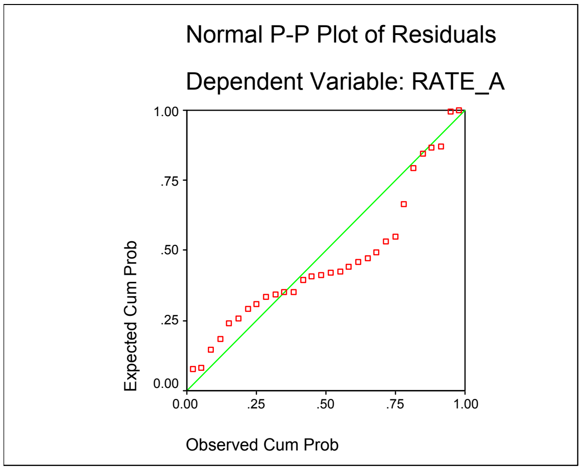
a Dependent Variable: RATE_B

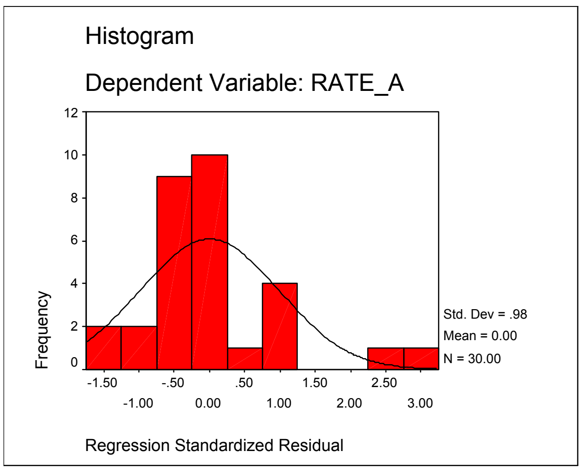
One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
Kolmogorov-Smirnov Z		.900
Asymp. Sig. (2-tailed)		.392

a Test distribution is Normal.
b Calculated from data.

Appendix E. Army Class A Residual Frequency Distribution and Normality Test





Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1.0743	3.4370	2.2557	.7172	30
Residual	-.8232	1.7195	-3.4491E-15	.5712	30

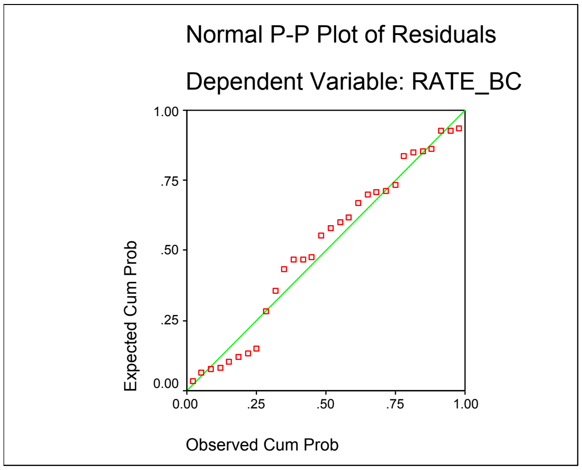
a Dependent Variable: RATE_A

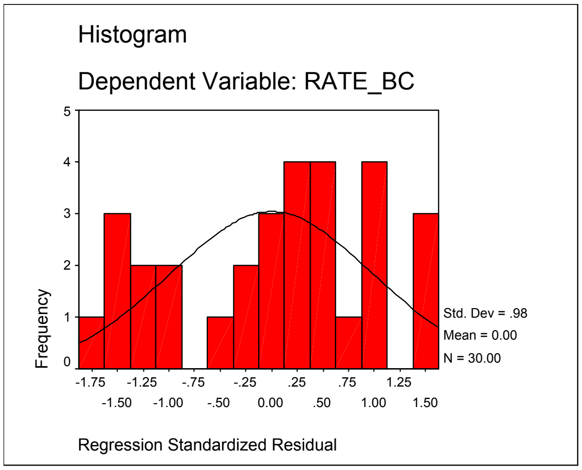
One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
Kolmogorov-Smirnov Z		1.192
Asymp. Sig. (2-tailed)		.117

a Test distribution is Normal.
b Calculated from data.

Appendix F. Army Class B-C Residual Frequency Distribution and Normality Test





Residuals Statistics

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	4.5415	16.2445	10.3930	3.5526	30
Residual	-7.4312	6.0939	4.086E-14	3.9610	30

a Dependent Variable: RATE_BC

One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
Kolmogorov-Smirnov Z		.656
Asymp. Sig. (2-tailed)		.782

a Test distribution is Normal.
b Calculated from data.

Appendix G. AF PPI Values

Class A					Class B			
FY	Rate	PPI	(x-mean)2		FY	Rate	PPI	(x-mean)^2
1983	1.73	10.00	0.088		1983	0.50	10.00	8.820
1984	1.77	10.22	0.006		1984	0.64	12.79	0.031
1985	1.49	8.41	3.540		1985	0.72	11.22	3.072
1986	1.79	12.05	3.074		1986	0.46	6.47	42.276
1987	1.51	8.41	3.561		1987	0.60	13.03	0.004
1988	1.64	10.89	0.357		1988	0.72	11.88	1.179
1989	1.59	11.00	0.495		1989	0.12	11.00	3.880
1990	1.49	9.37	0.858		1990	0.39	32.89	396.727
1991	1.11	7.49	7.869		1991	0.43	11.24	2.980
1992	1.69	15.15	23.572		1992	0.39	9.09	15.078
1993	1.35	7.98	5.345		1993	0.59	15.05	4.329
1994	1.46	10.87	0.325		1994	0.71	11.94	1.057
1995	1.44	12.00	2.903	Pre-ORM	1995	0.86	12.00	0.941
1996	1.24	8.61	4.175	Post-ORM	1996	0.51	5.91	95.790
1997	1.37	10.99	0.113		1997	0.71	13.96	3.029
1998	1.14	8.31	5.517		1998	0.43	6.02	93.590
1999	1.55	13.62	8.786		1999	1.22	28.62	166.978
2000	1.08	6.98	13.566		2000	4.08	33.40	313.329
2001	1.16	13.00	5.484		2001	3.68	13.00	7.282
2002	1.52	13.09	5.922		2002	3.30	8.98	45.197
83-95		96-02			83-95		96-02	
mean =	10.30	mean =	10.66		mean =	12.97	mean =	15.70
variance =	4.33	variance =	17.19		variance =	40.03	variance =	120.87
	Sp^2 =	5.06	---pooled variance---			Sp^2 =	65.23	

Appendix H. Army PPI Values

Class A					Class B-C			
FY	Rate	PPI	(x-mean)2		FY	Rate	PPI	(x-mean)2
1973	4.09	10.00	0.008		1987	18.02	10.00	0.250
1974	3.24	7.92	3.954		1987	16.85	9.35	0.022
1975	3.52	10.86	0.910		1987	16.65	9.88	0.143
1976	3.24	9.20	0.498		1987	19.41	11.66	4.669
1977	3	9.26	0.424		1987	18.81	9.69	0.037
1978	3.1	10.33	0.179		1987	16.76	8.91	0.349
1979	2.7	11.00	1.187		1987	16.07	9.59	0.008
1980	2.34	8.67	1.547		1987	19.25	11.98	6.159
1981	2.63	11.24	1.766		1987	19.11	9.93	0.181
1982	3.29	12.51	6.756		1987	11.14	5.83	13.474
1983	2.33	7.08	7.999		1987	8.05	7.23	5.153
1984	2.53	10.86	0.899		1987	5.65	7.02	6.138
1985	2.94	12.00	4.367		1987	5.61	9.93	0.184
1986	1.97	6.70	10.302		1987	5.90	10.50	1.005
1987	2.17	11.02	1.221	Pre-RM	1987	4.93	11.00	2.251
1988	1.84	8.48	6.103	Post-RM	1988	2.76	5.59	28.908
1989	1.90	10.33	0.389		1989	5.28	19.17	67.228
1990	1.83	9.63	1.737		1990	4.67	8.85	4.498
1991	3.69	13.00	4.204		1991	8.31	17.78	46.426
1992	1.57	4.25	44.823		1992	6.29	7.56	11.588
1993	1.77	11.27	0.105		1993	7.93	12.61	2.701
1994	1.64	9.27	2.836		1994	7.43	9.38	2.534
1995	0.83	5.06	34.677		1995	7.06	9.50	2.156
1996	0.74	8.92	4.137		1996	7.49	10.60	0.135
1997	1.26	14.00	9.304		1997	7.35	9.81	1.338
1998	1.34	10.63	0.099		1998	8.24	11.22	0.064
1999	1.97	14.70	14.076		1999	11.60	14.07	9.649
2000	0.62	3.14	60.958		2000	7.54	6.50	19.967
2001	1.02	16.51	30.922		2001	8.87	12.00	1.064
2002	2.56	25.05	198.788		2002	8.76	9.88	1.193
83-95		96-02			83-95		96-02	
mean =	9.91	mean =	10.95		mean =	9.50	mean =	10.97
variance =	3.00	variance =	29.51		variance =	2.86	variance =	14.25
	Sp^2 =	15.98		---pooled variance---		Sp^2 =	8.82	

Appendix I. AF Exponential Smoothing Transformation

AF Class A				AF Class B			
Year	Observation	Smoothed Value	Smoothed Trend	Year	Observation	Smoothed Value	Smoothed Trend
72	3.04	3.04	0.00	72	0.97	3.04	0.00
73	2.37	2.84	-0.06	73	0.98	2.42	-0.19
74	2.89	2.81	-0.05	74	0.88	1.83	-0.31
75	2.77	2.76	-0.05	75	0.68	1.27	-0.38
76	2.81	2.74	-0.04	76	0.68	0.83	-0.40
77	2.84	2.74	-0.03	77	9.45	3.13	0.41
78	3.16	2.85	0.01	78	13.02	6.39	1.26
79	2.95	2.89	0.02	79	2.10	5.98	0.76
80	2.56	2.80	-0.01	80	1.80	5.26	0.32
81	2.44	2.69	-0.04	81	1.67	4.41	-0.03
82	2.33	2.55	-0.07	82	0.48	3.20	-0.38
83	1.73	2.25	-0.14	83	0.50	2.12	-0.59
84	1.77	2.01	-0.17	84	0.64	1.26	-0.67
85	1.49	1.74	-0.20	85	0.72	0.63	-0.66
86	1.79	1.61	-0.18	86	0.46	0.11	-0.62
87	1.51	1.46	-0.17	87	0.60	-0.17	-0.52
88	1.64	1.39	-0.14	88	0.72	-0.27	-0.39
89	1.59	1.35	-0.11	89	0.12	-0.43	-0.32
90	1.49	1.32	-0.09	90	0.39	-0.41	-0.22
91	1.11	1.19	-0.10	91	0.43	-0.31	-0.12
92	1.69	1.27	-0.04	92	0.39	-0.18	-0.05
93	1.35	1.26	-0.03	93	0.59	0.01	0.02
94	1.46	1.30	-0.01	94	0.71	0.24	0.09
95	1.44	1.33	0.00	95	0.86	0.49	0.13
96	1.24	1.31	-0.01	96	0.51	0.58	0.12
97	1.37	1.32	0.00	97	0.71	0.71	0.12
98	1.14	1.26	-0.02	98	0.43	0.71	0.09
99	1.55	1.34	0.01	99	1.22	0.92	0.12
100	1.08	1.27	-0.01	100	4.08	1.96	0.40
101	1.16	1.23	-0.02	101	3.68	2.75	0.52
102	1.52	1.30	0.01	102	3.30	3.28	0.52
alpha	0.3			alpha	0.3		
beta	0.3			beta	0.3		
MAD	alpha/beta			MAD	alpha/beta		
3.64	0.5			26.41	0.5		
5.55	0.3			40.63	0.3		
9.44	0.1			55.97	0.1		

Appendix J. Army Exponential Smoothing Transformation

Army Class A				Army Class B-C			
Year	Observation	Smoothed Value	Smoothed Trend	Year	Observation	Smoothed Value	Smoothed Trend
73	4.09	4.09	0.00	73	18.02	18.02	0.00
74	3.24	3.84	-0.08	74	16.85	17.67	-0.10
75	3.52	3.69	-0.10	75	16.65	17.29	-0.19
76	3.24	3.48	-0.13	76	19.41	17.80	0.02
77	3.00	3.25	-0.16	77	18.81	18.12	0.11
78	3.10	3.09	-0.16	78	16.76	17.79	-0.02
79	2.70	2.86	-0.18	79	16.07	17.26	-0.17
80	2.34	2.58	-0.21	80	19.25	17.73	0.02
81	2.63	2.45	-0.19	81	19.11	18.16	0.14
82	3.29	2.57	-0.10	82	11.14	16.15	-0.50
83	2.33	2.43	-0.11	83	8.05	13.37	-1.19
84	2.53	2.38	-0.09	84	5.65	10.23	-1.77
85	2.94	2.49	-0.03	85	5.61	7.60	-2.03
86	1.97	2.31	-0.07	86	5.90	5.67	-2.00
87	2.17	2.22	-0.08	87	4.93	4.05	-1.89
88	1.84	2.05	-0.11	88	2.76	2.34	-1.83
89	1.90	1.93	-0.11	89	5.28	1.94	-1.40
90	1.83	1.82	-0.11	90	4.67	1.78	-1.03
91	3.69	2.30	0.07	91	8.31	3.01	-0.35
92	1.57	2.13	0.00	92	6.29	3.75	-0.02
93	1.77	2.02	-0.04	93	7.93	4.99	0.35
94	1.64	1.88	-0.07	94	7.43	5.97	0.54
95	0.83	1.52	-0.16	95	7.06	6.68	0.59
96	0.74	1.18	-0.21	96	7.49	7.33	0.61
97	1.26	1.05	-0.19	97	7.35	7.76	0.56
98	1.34	1.01	-0.14	98	8.24	8.30	0.55
99	1.97	1.20	-0.04	99	11.60	9.67	0.80
100	0.619	0.99	-0.09	100	7.54	9.59	0.53
101	1.022	0.94	-0.08	101	8.87	9.75	0.42
102	2.56	1.37	0.07	102	8.76	9.75	0.29
alpha	0.3			alpha	0.3		
beta	0.3			beta	0.3		
MAD	alpha/beta			MAD	alpha/beta		
7.65	0.5			32.77	0.5		
9.79	0.3			55.33	0.3		
15.47	0.1			104.15	0.1		

Appendix K. AF Comparison of Means Tests, Rates

1. Simple comparison of means

pre_post		A_AF	B_AF
pre_orm	Mean	1.5431	.5485
	N	13	13
	Std. Deviation	.1892	.1942
post_orm	Mean	1.2943	1.9900
	N	7	7
	Std. Deviation	.1883	1.6225
Total	Mean	1.4560	1.0530
	N	20	20
	Std. Deviation	.2205	1.1631

Class A decrease.
Class B increase.

2. ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
RATE_A	Between Groups	.282	1	.282	7.891	.012
	Within Groups	.642	18	3.569E-02		
	Total	.924	19			
RATE_B	Between Groups	9.455	1	9.455	10.474	.005
	Within Groups	16.248	18	.903		
	Total	25.703	19			

Class A significance less than 0.05, so reject null hypothesis—means are not equal.
Class B significance less than 0.05, so reject null hypothesis—means are not equal.

Appendix K. AF Comparison of Means Tests, Rates, continued

3. T-Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
RATE_A	Equal variances assumed	.072	.792	2.809	18	.012	.2488	8.857E-02	6.272E-02	.4349
	Equal variances not assumed			2.813	12.457	.015	.2488	8.843E-02	5.689E-02	.4407
RATE_B	Equal variances assumed	121.947	.000	-3.236	18	.005	-1.4415	.4454	-2.3773	-.5058
	Equal variances not assumed			-2.342	6.093	.057	-1.4415	.6156	-2.9424	5.929E-02

Class A significance less than 0.025, so reject null hypothesis—means are not equal.
Class B significance less than 0.025 when equal variances assumed—so reject null hypothesis—means are not equal. Do not reject when equal variances are not assumed.

4. Mann Whitney U and Wilcoxon W tests

	RATE_A	RATE_B
Mann-Whitney U	19.000	20.000
Wilcoxon W	47.000	111.000
Z	-2.101	-2.024
Asymp. Sig. (2-tailed)	.036	.043
Exact Sig. [2*(1-tailed Sig.)]	.037	.046

Class A significance greater than 0.025, so do not reject null hypothesis—means are equal.
Class B significance greater than 0.025, so do not reject null hypothesis—means are equal.

Appendix L. Army Comparison of Means Tests, Rates

1. Simple comparison of means

rm		RATE_A	RATE_BC
pre rm	Mean	2.8727	13.4807
	N	15	15
	Std. Deviation	.5676	5.8359
post rm	Mean	1.6387	7.3053
	N	15	15
	Std. Deviation	.7768	2.0387
Total	Mean	2.2557	10.3930
	N	30	30
	Std. Deviation	.9169	5.3208

Class A decrease.
Class B/C decrease.

2. ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
RATE_A	Between Groups	11.421	1	11.421	24.676	.000
	Within Groups	12.959	28	.463		
	Total	24.380	29			
RATE_BC	Between Groups	286.011	1	286.011	14.969	.001
	Within Groups	534.999	28	19.107		
	Total	821.010	29			

Class A significance less than 0.05, so reject null hypothesis—means are not equal.
Class B/C significance less than 0.05, so reject null hypothesis—means are not equal.

3. T-Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
RATE_A	Equal variances assumed	.297	.590	4.968	28	.000	1.2340	.2484	.7251	1.7429
	Equal variances not assumed			4.968	25.633	.000	1.2340	.2484	.7230	1.7450
RATE_BC	Equal variances assumed	36.213	.000	3.869	28	.001	6.1753	1.5961	2.9058	9.4448
	Equal variances not assumed			3.869	17.367	.001	6.1753	1.5961	2.8132	9.5375

Class A significance less than 0.025, so reject null hypothesis—means are not equal.

Class B/C significance less than 0.025, so reject null hypothesis—means are not equal.

4. Mann Whitney U and Wilcoxon W tests

	RATE_A	RATE_BC
Mann-Whitney U	19.500	55.000
Wilcoxon W	139.500	175.000
Z	-3.858	-2.385
Asymp. Sig. (2-tailed)	.000	.017
Exact Sig. [2*(1-tailed Sig.)]	.000	.016

Class A significance less than 0.025, so reject null hypothesis—means are not equal.
Class B/C significance less than 0.025, so reject null hypothesis—means are not equal.

Appendix M. AF Comparison of Means Tests, PPI

1. Simple comparison of means

ORM		PPI_A	PPI_B
Pre ORM	Mean	10.2954	12.9692
	N	13	13
	Std. Deviation	2.0821	6.3273
Post ORM	Mean	10.6571	15.6986
	N	7	7
	Std. Deviation	2.6932	10.9942
Total	Mean	10.4220	13.9245
	N	20	20
	Std. Deviation	2.2494	8.0771

Class A slight increase.
Class B increase.

2. ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
PPI_A	Between Groups	.595	1	.595	.112	.742
	Within Groups	95.540	18	5.308		
	Total	96.136	19			
PPI_B	Between Groups	33.894	1	33.894	.506	.486
	Within Groups	1205.659	18	66.981		
	Total	1239.553	19			

Class A significance greater than 0.05, so do not reject null hypothesis—means are equal.
Class B significance greater than 0.05, so do not reject null hypothesis—means are equal.

Appendix M. AF Comparison of Means Tests, PPI, continued

3. T-Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
PPI_A	Equal variances assumed	1.695	.209	-.335	18	.742	-.3618	1.0801	-2.6309	1.9074
	Equal variances not assumed			-.309	9.967	.764	-.3618	1.1703	-2.9705	2.2470
PPI_B	Equal variances assumed	4.517	.048	-.711	18	.486	-2.7293	3.8368	-10.7902	5.3315
	Equal variances not assumed			-.605	8.201	.561	-2.7293	4.5108	-13.0871	7.6284

Class A significance greater than 0.025, so do not reject null hypothesis—means are equal.
Class B significance greater than 0.025, so do not reject null hypothesis—means are equal.

4. Mann Whitney U and Wilcoxon W tests

	PPI_A	PPI_B
Mann-Whitney U	40.000	44.000
Wilcoxon W	131.000	135.000
Z	-.436	-.119
Asymp. Sig. (2-tailed)	.663	.905
Exact Sig. [2*(1-tailed Sig.)]	.699	.938

Class A significance greater than 0.025, so do not reject null hypothesis—means are equal.
Class B significance greater than 0.025, so do not reject null hypothesis—means are equal.

Appendix N. Army Comparison of Means Tests, PPI

1. Simple means comparison.

rm		PPI_A	PPI_BC
pre rm	Mean	9.9100	9.5000
	N	15	15
	Std. Deviation	1.7330	1.6907
post rm	Mean	10.9493	10.9680
	N	15	15
	Std. Deviation	5.4330	3.7744
Total	Mean	10.4297	10.2340
	N	30	30
	Std. Deviation	3.9974	2.9690

Slight increases in both PPIs.

2. ANOVA.

		Sum of Squares	df	Mean Square	F	Sig.
PPI_A	Between Groups	8.102	1	8.102	.498	.486
	Within Groups	455.289	28	16.260		
	Total	463.391	29			
PPI_BC	Between Groups	16.163	1	16.163	1.890	.180
	Within Groups	239.464	28	8.552		
	Total	255.626	29			

Class A significance greater than 0.05, so do not reject null hypothesis—means are equal.
Class B/C significance greater than 0.05, so do not reject null hypothesis—means are equal.

3. Independent T-test.

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
PPI_A	Equal variances assumed	6.081	.020	-.706	28	.486	-1.0393	1.4724	-4.0555	1.9768
	Equal variances not assumed			-.706	16.820	.490	-1.0393	1.4724	-4.1484	2.0698
PPI_BC	Equal variances assumed	5.323	.029	-1.375	28	.180	-1.4680	1.0678	-3.6554	.7194
	Equal variances not assumed			-1.375	19.401	.185	-1.4680	1.0678	-3.6999	.7639

Class A significance greater than 0.025, so do not reject null hypothesis—means are equal.
Class B/C significance greater than 0.025, so do not reject null hypothesis—means are equal.

4. Mann Whitney

	PPI_A	PPI_BC
Mann-Whitney U	102.500	93.500
Wilcoxon W	222.500	213.500
Z	-.415	-.788
Asymp. Sig. (2-tailed)	.678	.431
Exact Sig. [2*(1-tailed Sig.)]	.683	.436

Class A significance greater than 0.025, so do not reject null hypothesis—means are equal.
Class B/C significance greater than 0.025, so do not reject null hypothesis—means are equal.

Appendix O. AF Comparison of Means, Exponential Smoothing

1. Simple means comparison.

orm		AF_A	AF_B
pre orm	Mean	2.0192	1.5996
	N	24	24
	Std. Deviation	.6897	2.1013
post orm	Mean	1.2867	1.7217
	N	6	6
	Std. Deviation	4.082E-02	1.1164
Total	Mean	1.8727	1.6240
	N	30	30
	Std. Deviation	.6829	1.9285

Class A decreased.
Class B increased.

2. ANOVA.

		Sum of Squares	df	Mean Square	F	Sig.
AF_A	Between Groups	2.575	1	2.575	6.586	.016
	Within Groups	10.950	28	.391		
	Total	13.525	29			
AF_B	Between Groups	7.154E-02	1	7.154E-02	.019	.893
	Within Groups	107.785	28	3.849		
	Total	107.857	29			

Class A significance less than 0.05, so reject—means are not equal.
Class B significance greater than 0.05, so do not reject—means are equal.

Appendix O. AF Comparison of Means, Exponential Smoothing, continued

3. Independent T-Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
AF_A	Equal variances assumed	48.044	.000	2.566	28	.016	.7325	.2854	.1478	1.3172
	Equal variances not assumed			5.167	23.628	.000	.7325	.1418	.4397	1.0253
AF_B	Equal variances assumed	2.280	.142	-.136	28	.893	-.1221	.8955	-1.9565	1.7123
	Equal variances not assumed			-.195	15.188	.848	-.1221	.6259	-1.4547	1.2105

Class A significance less than 0.025, so reject null hypothesis—means are not equal.
Class B significance greater than 0.025, so do not reject null hypothesis—means are equal.

4. Mann-Whitney

	AF_A	AF_B
Mann-Whitney U	21.000	53.000
Wilcoxon W	42.000	353.000
Z	-2.646	-.985
Asymp. Sig. (2-tailed)	.008	.325
Exact Sig. [2*(1-tailed Sig.)]	.006	.347

Class A significance less than 0.025, so reject null hypothesis—means are not equal.
Class B significance greater than 0.025, so do not reject null hypothesis—means are equal.

Appendix P. Army Comparison of Means, Exponential Smoothing

1. Simple means comparison.

rm		AR_A	AR_BC
pre orm	Mean	2.9153	14.4607
	N	15	15
	Std. Deviation	.6153	5.0248
post orm	Mean	1.5593	6.1740
	N	15	15
	Std. Deviation	.4792	3.0161
Total	Mean	2.2373	10.3173
	N	30	30
	Std. Deviation	.8770	5.8600

Class A decreased.
Class B/C decreased.

2. ANOVA.

		Sum of Squares	df	Mean Square	F	Sig.
AR_A	Between Groups	13.791	1	13.791	45.346	.000
	Within Groups	8.515	28	.304		
	Total	22.306	29			
AR_BC	Between Groups	515.016	1	515.016	29.990	.000
	Within Groups	480.836	28	17.173		
	Total	995.853	29			

Class A significance less than 0.05, so reject—means are not equal.
Class BC significance less than 0.05, so reject—means are not equal.

Appendix P. Army Comparison of Means, Exponential Smoothing, continued

3. Independent T-Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
AR_A	Equal variances assumed	1.252	.273	6.734	28	.000	1.3560	.2014	.9435	1.7685
	Equal variances not assumed			6.734	26.415	.000	1.3560	.2014	.9424	1.7696
AR_BC	Equal variances assumed	4.536	.042	5.476	28	.000	8.2867	1.5132	5.1871	11.3863
	Equal variances not assumed			5.476	22.929	.000	8.2867	1.5132	5.1559	11.4174

Class A significance less than 0.025, so reject null hypothesis—means are not equal.
Class B/C significance less than 0.025, so reject null hypothesis—means are not equal.

4. Mann-Whitney

	AR_A	AR_BC
Mann-Whitney U	1.000	25.000
Wilcoxon W	121.000	145.000
Z	-4.625	-3.630
Asymp. Sig. (2-tailed)	.000	.000
Exact Sig. [2*(1-tailed Sig.)]	.000	.000

Class A significance less than 0.025, so reject null hypothesis—means are not equal.
Class B/C significance less than 0.025, so reject null hypothesis—means are not equal.

Appendix Q. AF Comparison of Variance

Class A				Class B-C		
FY	Rate	Residuals		FY	Rate	Residuals
1983	1.73	0.04457		1983	0.5	0.65643
1984	1.77	0.10872		1984	0.64	0.66912
1985	1.49	-0.14713		1985	0.72	0.62181
1986	1.79	0.17702		1986	0.46	0.2345
1987	1.51	-0.07883		1987	0.6	0.2472
1988	1.64	0.07532		1988	0.72	0.23989
1989	1.59	0.04947		1989	0.12	-0.48742
1990	1.49	-0.02638		1990	0.39	-0.34473
1991	1.11	-0.38223		1991	0.43	-0.43204
1992	1.69	0.22192		1992	0.39	-0.59935
1993	1.35	-0.09392		1993	0.59	-0.52665
1994	1.46	0.04023		1994	0.71	-0.53396
1995	1.44	0.04438		1995	0.86	-0.51127
1996	1.24	-0.13147		1996	0.51	-0.98858
1997	1.37	0.02268		1997	0.71	-0.91589
1998	1.14	-0.18317		1998	0.43	-1.3232
1999	1.55	0.25098		1999	1.22	-0.6605
2000	1.08	-0.19487		2000	4.08	2.07219
2001	1.16	-0.09072		2001	3.68	1.54488
2002	1.52	0.29343		2002	3.3	1.03757
	s^2_pre	0.024124			s^2_pre	0.259991
	n_pre	14			n_pre	14
	df_pre	13			df_pre	13
	s^2_post	0.041082			s^2_post	2.053957
	n_post	7			n_post	7
	df_post	6			df_post	6
	F-stat	1.702905			F-stat	7.900109
	F-crit	2.92			F-crit	2.92
	variances are equal				variances are not equal	

Appendix R. Army Comparison of Variance

Class A				Class B-C		
FY	Rate	Residuals		FY	Rate	Residuals
1973	4.09	0.65301		1973	18.02	1.77555
1974	3.24	-0.11552		1974	16.85	1.0091
1975	3.52	0.24595		1975	16.65	1.21265
1976	3.24	0.04742		1976	19.41	4.37619
1977	3	-0.11111		1977	18.81	4.17974
1978	3.1	0.07036		1978	16.76	2.53329
1979	2.7	-0.24817		1979	16.07	2.24684
1980	2.34	-0.5267		1980	19.25	5.83039
1981	2.63	-0.15523		1981	19.11	6.09394
1982	3.29	0.58625		1982	11.14	-1.47252
1983	2.33	-0.29228		1983	8.05	-4.15897
1984	2.53	-0.01081		1984	5.65	-6.15542
1985	2.94	0.48066		1985	5.61	-5.79187
1986	1.97	-0.40787		1986	5.9	-5.09832
1987	2.17	-0.1264		1987	4.93	-5.66477
1988	1.84	-0.37493		1988	2.76	-7.43123
1989	1.9	-0.23346		1989	5.28	-4.50768
1990	1.83	-0.22199		1990	4.67	-4.71413
1991	3.69	1.71948		1991	8.31	-0.67058
1992	1.57	-0.31905		1992	6.29	-2.28703
1993	1.77	-0.03758		1993	7.93	-0.24348
1994	1.64	-0.08611		1994	7.43	-0.33994
1995	0.83	-0.81464		1995	7.06	-0.30639
1996	0.74	-0.82317		1996	7.49	0.52716
1997	1.26	-0.2217		1997	7.35	0.79071
1998	1.34	-0.06023		1998	8.24	2.08426
1999	1.97	0.65124		1999	11.6	5.84781
2000	0.62	-0.61729		2000	7.54	2.19135
2001	1.02	-0.13581		2001	8.87	3.9249
2002	2.56	1.48566		2002	8.76	4.21845
	s^2_pre	0.131246			s^2_pre	18.35831
	n_pre	14			n_pre	14
	df_pre	13			df_pre	13
	s^2_post	0.516631			s^2_post	14.03625
	n_post	16			n_post	16
	df_post	15			df_post	15
	F-stat	3.936			F-stat	0.765
	F-crit	2.530			F-crit	2.480
	variances are not equal				variances are equal	

Appendix S. Human Factors Proportions Test Results

AF Class A				
sum[(f-e)^2]/2	2787.57		Number Increased	6
df	15		Number Decreased	6
crit	32.80		Number Unchanged	4
Reject Null Hypothesis				

AF Class A--Test Two				
sum[(f-e)^2]/2	110.16		Number Increased	4
df	10		Number Decreased	6
crit	18.31		Number Unchanged	1
Reject Null Hypothesis				

AF Class B				
sum[(f-e)^2]/2	379.41		Number Increased	5
df	15		Number Decreased	8
crit	32.80		Number Unchanged	3
Reject Null Hypothesis				

Army Class A				
sum[(f-e)^2]/2	111.46		Number Increased	5
df	12		Number Decreased	4
crit	28.29		Number Unchanged	4
Reject Null Hypothesis				

Army Class B				
sum[(f-e)^2]/2	372.29		Number Increased	13
df	12		Number Decreased	0
crit	28.29		Number Unchanged	0
Reject Null Hypothesis				

Bibliography

Air Force Safety Center. "Air Force Safety Analysis." Briefing Slides, n. pag. http:// http://safety.kirtland.af.mil/AFSC/files/tome2.pdf. 15 February 2003a.

Air Force Safety Center. Aviation Mishap Data. n. pag. http://safety.kirtland.af.mil/AFSC/RDBMS/Flight/stats/usaf1097.html. November 2002.

Air Force Safety Center. Protected Aviation Mishap Data. January 2003b.

Ashley, Park D. *Operational Risk Management and Military Aviation Safety.* MS thesis, AFIT/GLM/LAL/99S-2. School of Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1999

"BASH." Excerpt from unpublished article. n. pag. http://safety.kirtland.af.mil/AFSC/Bash/home.html. 26 September 2002.

Brandon, Linda. "Operations Tempo Tied to Fatal Helicopter Crash." Excerpt from unpublished article. n. pag. http://www.af.mil/news/Mar1999/n19990316_990412.html. October 2002.

Army Safety Center. Aviation Mishap Data. n. pag. http://safety.army.mil. November 2002.

"Aviation Studies." Excerpt from unpublished article. n. pag. http://www.nasdac.faa.gov/aviation_studies/weather_study. November 2002.

"Birdstike Committee USA." Excerpt from unpublished article. n. pag. http://www.birdstrike.org. 26 September 2002.

Bowerman, B. L. and O'Connell, J. C. *Time Series Forecasting.* Boston: Duxbury Press, 1987.

Cantu, R. *The Role of Weather in Major Naval Aviation Mishaps FY 90-98*. MS thesis. Naval Postgraduate School, Monterey, CA, March 2001. (AD-A391038)

Castro, C.A. and A. B. Adler. "OPTEMPO: Effects on Soldier and Unit Readiness." *Parameters*. 29: 86-95 (Autumn 1999).

Dahlman, C., R. Kerchner, and D. Thaler. *Setting Requirements for Maintenance Manpower in the US Air Force*. Santa Monica California: RAND, 2002.

Department of the Air Force. *The Blue Ribbon Panel on Aviation Safety*. Washington: HQ USAF. 5 September 1995.

Department of the Air Force. *Operational Risk Management*. AFI 90-901. Washington: HQ USAF, 1 April 2000a.

Department of the Air Force. *Operational Risk Management*. AFPD 90-9. Washington: HQ USAF, 1 April 2000b.

Department of the Air Force. *Operational Risk Management Guidelines and Tools*. AFPAM 90-902. Washington: HQ USAF, 14 December 2000c.

Department of the Air Force. *Safety Investigations and Reports*. AFI 91-204. Washington: HQ USAF, 11 December 2001.

Department of the Army. *Army Aviation Accident Prevention*. AR 385-95. Washington: HQ US Army, 10 December 1999.

Department of the Army. *Army Accident Investigating and Reporting*. DAPAM 385-40. Washington: HQ US Army, 1 November 1994.

Department of the Army. *Risk Management*. FM 100-14. Washington: HQ US Army, 23 April 1998.

Department of Defense. *Accident Investigation, Reporting, and Record Keeping*. DODI 6055.7. Washington: Pentagon, 3 October 2000.

Department of Defense. *Report of the Defense Science Board Task Force on Aviation Safety*. Washington, February, 1997.

Driskell, James E. and Richard J. Adams. *Crew Resource Management: An Introductory Handbook*. Washington: Department of Transportation, August 1992.

Duquette, Alison. "Fact Sheet: Aviation Accident Statistics." Excerpt from unpublished article. n. pag. www.faa.gov/apa/safer-skies/fsstats.htm. 26 September 2001.

Fitzsimmons, James A. and Mona J. Fitzsimmons. *Service Management*. New York: McGraw-Hill, 2001.

"Human Factors." Excerpt from unpublished article. n. pag. http://human-factors.arc.nasa.gov/zteam. November 2002.

Johnson, C. "*Reasons for the Failure of CRM Training in Aviation*." Excerpt from unpublished article. n. pag. http//www.dcs.gla.ac.uk. November 2002.

Jumper, John J. Air Force Chief of Staff, Department of the Air Force, Washington DC. Memorandum on Operational Risk Management. 26 Jun 2002a.

Jumper, John J. Air Force Chief of Staff, Department of the Air Force, Washington DC. Memorandum on Operational Risk Management. 20 Dec 2002b.

Leedy, P. D. and J. E. Ormrod. *Practical Research.* New Jersey: Prentice Hall, Inc. 2001.

The Merriam Webster Dictionary. Springfield: Merriam-Webster, Incorporated, 1994.

National Security and International Affairs Division*. Military Aircraft Safety: Significant Improvements Since 1975*. Washington: General Accounting Office, 1 February 1996.

Neter, John, Michael H. Kutner, Christopher J. Nachtsheim and William Wasserman. *Applied Linear Statistical Models*. 4th edition. Burr Ridge IL: Irwin, 1996.

Rebok, G.W., G. Li., S. P. Baker, J. G. Grabowski, and S. Willoughby. "Self-rated Changes in Cognition and Piloting Skills: a Comparison of Younger and Older Airline Pilots." *Aviation, Space, and Environmental Medicine*, 73: 466-471 (2002).

Salas, E., C.S. Burke, C. A. Bowers, and K. A. Wilson. "Team Training in the Skies: Does Crew Resource Management (CRM) Training Work?" *Human Factors*, 43: 641-674 (2001).

Schilder, C. "Accident Investigation and Analysis." Excerpt from unpublished article. n. pag. http://www.denix.osd.mil/denix. September 2002.

Shappell, S.A. and D.A. Wiegmann. *The Human Factors Analysis and Classification System-HFACS*. Report No. DOT/FAA/AM-00/7. Washington: Office of Aviation Medicine, February 2000.

Shappel, S.A. and D.A. Wiegmann. "Unraveling the Mystery of General Aviation Controlled Flight Into Terrain Accidents Using HFAC." *A paper presented at the 11th International Symposium on Aviation Psychology.* The Ohio State University, Columbus OH: 2001.

"Status of the United States Military." Except from unpublished article. n. pag. http://www.ndia.org/dvocacy/resources/hearings. 2 October 2002.

Weigmann, D.A. and S. A. Shappell. "Human Error and Crew Resource Management Failures in Naval Aviation Mishaps: A Review of U.S. Naval Safety Center Data, 1990-1996." *ASME*, 70: 1147-51 (1999).

Vita

Captain Matthew G. Cho graduated from Moanalua High School in Honolulu, Hawaii. He entered undergraduate studies at the University of Kansas in Lawrence, Kansas where he graduated with honors with a Bachelor of Architecture degree in September 1998. He was commissioned through the Detachment 280 AFROTC at the University of Kansas where he was recognized as a Distinguished Graduate.

His first assignment was at Hill AFB as the 388^{th} FW Plans and Programs officer. In Feb 1999, he was assigned to the 729^{th} Air Control Squadron, Hill AFB, Utah where he served as the Combat Support Director and Squadron Mobility Officer. While stationed at Hill, he attended the Logistics Plans Officer School at Lackland AFB, Texas where he graduated as a Distinguished Graduate. In October 2001, he was assigned to the 51^{st} Logistics Support Squadron at Osan AB, Republic of Korea and served as the 51^{st} FW War Reserve Materiel Officer and alternate Installation Deployment Officer. In August 2002, he entered the Graduate School of Engineering and Management, Air Force Institute of Technology. Upon graduation, he will be assigned to the C-17 Systems Program Office at Wright Patterson AFB, Ohio.

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 074-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
14-03-03	**Master's Thesis**	Sept 02 – Mar 03

4. TITLE AND SUBTITLE	
THE AIR FORCE OPERATIONAL RISK MANAGEMENT PROGRAM AND AVIATION SAFETY	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Cho, Matthew G., Captain, USAF	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 P Street, Building 640 WPAFB OH 45433-7765	AFIT/GLM/ENS/03-02
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
Air Force Safety Center Policy, Research, and Technology Division (AFSC/SEPR) HQ AFSC/SEPR 9700 G Ave SE, Bldg 24499 Kirtland AFB NM 87117-5670	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The Air Force implemented the Operational Risk Management (ORM) program in 1996 in an effort to protect their most valuable resources: aircraft and aviators. An AFIT thesis conducted in 1999 by Capt Park Ashley studied the Army's similar Risk Management (RM) program. Ashley concluded that since his analysis found that RM did not affect the Army's mishap rates, the AF should not expect to see its rates decline due to ORM implementation.

The purpose of this thesis was to determine whether the implementation of ORM has had any affect on the AF's mishap rates. Analysis was conducted on annual and quarterly mishap rates, quarterly sortie mishap rates, and individual mishap data using three statistical techniques: comparison of means testing, discontinuous piecewise linear regression, and chi-squared goodness of fit testing. Results showed that the implementation of ORM did not effectively reduce the Air Force's aviation mishap rates.

15. SUBJECT TERMS
Operational Risk Management, Safety, Aviation Mishaps, Accidents, Risk, Risk Management

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Stephen M. Swartz, Lt Col, USAF (ENS)
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	UU	159	(937) 255-6565, ext 4285; e-mail: Stephen.swartz@afit.edu